The Attrition of French as a Foreign Language

Bert Weltens

THE ATTRITION OF FRENCH AS A FOREIGN LANGUAGE

THE ATTRITION OF FRENCH

AS A FOREIGN LANGUAGE

een wetenschappelijke proeve op het gebied van de Letteren

Proefschrift
ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen
volgens het besluit van het college van decanen
in het openbaar te verdedigen op woensdag
14 december 1988, des namiddags te 3.30 uur

door

HUBERTUS HENRICUS GERTRUDA WELTENS

geboren op 23 juli 1957
te Heer-Maastricht

Dordrecht/Providence: Foris Publications

Promotor	Prof dr T J M van Els
Co-referent	Dr C L J de Bot

CONTENTS

PREFACE

The present book is the final report of a four-year research project that was funded by the Dutch Linguistics Foundation, which is sponsored by the Netherlands Organization for Scientific Research (NWO), under grant no. 300-167-006. The project consisted in an explorative investigation of the long-term retention of French foreign-language skills acquired by students in Dutch secondary schools.

The project was started in February 1984, and had originally been planned to finish in February 1988. Due to the fact that I was offered a part-time job as a university lecturer at the Department of Applied Linguistics of the University of Nijmegen in September 1987, the last five (full-time) months of the project were turned into twelve months on a 0.4 basis. As a result, the project ran until September 1988.

I would like to thank the following people for their assistance in the making of this dissertation: *Hubert Dartenay*, for his willingness to pronounce endless lists of stimuli exactly as I wanted them; *Wim Mulder* of the teacher training college *Interstudie*, for helping me with the choice of my cloze test; *The Dutch National Institute for Educational Measurement* (CITO), for allowing me to use one of their listening tests and for providing me with several pieces of useful information - I would like to single out *Trude Maas-de Brouwer*, who was then still at CITO, as having been particularly helpful, both in the process of obtaining information and services from CITO, and in the process of constructing the tests that I had to develop myself; *Elly Kersjes*, for typing the first version of my manuscript, and for assisting me in my struggle with the laser printers; *Leo Noordman* and *Rob Schreuder*, the former and present Director of the Interfaculty Research Unit for Language and Speech (IWTS) respectively, for helping me with various psycholinguistic aspects of my project, and for re-assuring me when I needed it; my student-assistants, *Marjon Grendel*, *Jos Poppe*, and *Paula Verkaik*, who took care of all kinds of laborious and tiresome jobs, and without whose knowledge of French I simply could not have carried out this investigation the way I have; and the staff at the two secondary schools I drew my

subjects from, in particular *Anneke Bouma*, *Harry Goossens*, *Annelies Kockelkoren*, and *Jan van Thiel*, who were always willing to help me out, often at very short notice. Finally, I would like to thank *Ralph B. Ginsberg* of the National Foreign Language Center, Washington DC, for various pieces of sound advice.

I also most heartily thank all my colleagues in the Faculty of Letters of the University of Nijmegen who offered assistance in various (essential) ways, but whom the regulations forbid me to name. I am quite sure they themselves know who are meant.

Nijmegen, August 1988 *Bert Weltens*

LIST OF ABBREVIATIONS

ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
CITO	Dutch National Institute for Educational Measurement
FL	foreign language
HI-COG	high-frequency cognates
HI-NOCOG	high-frequency non-cognates
LC	linguistically-competent speaker
LEX-LI	lexical listening (test)
LEX-RE	lexical reading (test)
LO-COG	low-frequency cognates
LO-NOCOG	low-frequency non-cognates
L1	mother tongue
L2	second language
MC	multiple-choice
MS-LI	morpho-syntactic listening (test)
MS-RE	morpho-syntactic reading (test)
PHO-LI	phonological listening (test)
PHO-RE	phonological reading (test)
SEC-*n*	grade *n* of secondary school
UNIV-*n*	year *n* of university

1. THE PHENOMENON OF FOREIGN-LANGUAGE ATTRITION

Foreign-language attrition, i.e. the deterioration of foreign-language skills, is a phenomenon that has only recently attracted some serious attention. Apart from some isolated efforts (Kennedy 1932; McMahon 1946; Geoghehan 1950; Scherer 1957; Flaugher & Spencer 1967; Pratella 1969, and Cohen 1974, 1975), it was not until May 1980, when a conference addressing the phenomenon was convened at the University of Pennsylvania (see Lambert & Freed 1982), that it became a source of real scientific interest. In that same year two dissertations were written on *second*-language (L2) attrition in young children (Allendorff 1980 and Hansen 1980), and one on *foreign*-language (FL) attrition in former FL students the next year (Godsall-Myers 1981). Since then, the field has expanded (see e.g. Lambert & Moore 1984, and Van Els 1986). Language attrition in its manifold manifestations (see below) has become one of the main research interests of the Department of Applied Linguistics of the University of Nijmegen, an expert symposium was gathered in the Netherlands to discuss the state of the art (with contributions from seven different countries; cf. Weltens et al. 1986a), and frequent contacts have been established between language attrition researchers all over the world.

Incidentally, the fact that we are dealing with a relatively recent field of research implies that we will in quite a number of instances be referring to material that has not (yet) been officially published, i.e. theses, internal reports, and the like.

What we have called the "field" of language attrition above, requires some definition. The most widely quoted definition of the term language attrition (or language loss) is that by Freed (1982:1) in her introductory article to Lambert & Freed (1982):

> "Broadly defined, language attrition may refer to the loss of any language or any portion of a language by an individual or a speech community".

Defined in this way, language attrition may refer to a variety of situations. De Bot & Weltens (1985) have classified language attrition research into

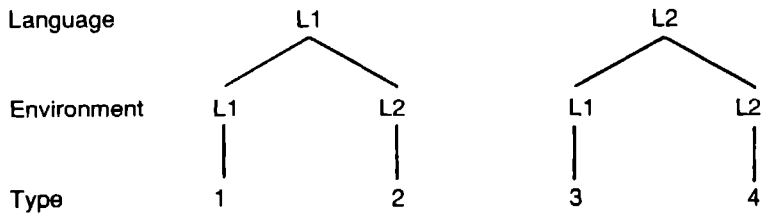

Fig. 1.1: Types of language attrition research.

four major categories, according to the nature of the language that is lost, an original mother tongue (L1) or an L2/FL, and according to the nature of the environment in which it is lost, an L1 or an L2 environment (see Figure 1.1). This classification is, as they themselves would readily admit, somewhat oversimplified, but it does provide a usable tool for distinguishing different types of research within the field. Examples of the four types are the following:

(1) the deterioration of language skills due to old age (Nyssen & Crahay 1960; Nicholas et al. 1982; De Bot & Lintsen 1986); the total extinction of a particular language or language variety: 'language death' (Dorian 1977, 1981; Dressler & Wodak-Leodolter 1977), and - when L1 is interpreted as a dialect - dialect loss *within* the dialect community (Schlieben-Lange 1977; Tabouret-Keller & Luckel 1981; Tsitsipis 1981; Hoppenbrouwers 1982; Trudgill 1983; Hagen & Münstermann 1985; Münstermann & Hagen 1986);

(2) the native languages of migrants (Campbell 1980; Clyne 1980; Jamieson 1980; Van Vlerken 1980; Galbraith 1981; Stankovski 1982; Stevens 1982; Appel 1983; Gonzo & Saltarelli 1983; Sharwood Smith 1983a, 1983b; Verheesen 1984; Boyd 1986; Davies 1986; Py 1986; Sawaie 1986; De Bot et al. 1988); the native languages of indigenous minorities (Hill & Hill 1977; Bauman 1980; Haugen et al. 1981; Fuller 1982), and dialect loss *outside* the dialect community (Daan 1969, 1971, 1987; Veldman 1975; Pauwels 1986);

(3) school-learned foreign languages - the subject of this book - and the deterioration of second languages after re-migration to the native country (Berman & Olshtain 1983; Aertssen et al. 1985; Olshtain 1986);

(4) the second languages of migrants, which seem to deteriorate with old age, resulting in L1 reversion (Clyne 1981; Hyltenstam & Stroud 1985).

Two cross-sections that are not represented as such in the figure are worth mentioning here. One, the distinction between 'natural' and 'pathological' loss, and, two, between *intra*generational and *inter*generational loss. 'Pathological' loss usually concerns the effects of brain damage of some sort, such as aphasia (see, e.g., Obler 1982), whereas 'natural' attrition refers to much more gradual and frequently occurring phenomena such as the displacement of one language (variety) by another, the processes occurring in languages-in-contact situations, and the deterioration of FL skills over longer periods of time in which the FL is not regularly used.

The dichotomy *intra-* versus *inter*generational attrition refers to processes *within individuals* and *across generations* respectively. In an L1 setting, for example, the two different processes could be characterised as 'mother tongue shift' and 'language displacement' respectively, although it is evident that attrition *across* generations of people probably also involves some attrition *within* each generation. It should be noted, however, that FL attrition, the object of the present research, is by definition an *intra*generational phenomenon, since it is *always* confined to individuals.

In some of the cases listed above, one should keep in mind that there are different phenomena that should be kept distinct from attrition. For dialect loss, for example, Mattheier (1986) distinguishes between processes such as the structural convergence of dialects towards the standard language ("Dialektverfall": dialect decay or dialect loss); the stuctural developments *within* the dialect system, occurring without any outside influence ("Dialektwandel": dialect change); and, finally, what Münstermann & Hagen (1986) have called "functional loss": the abandonment of the dialect in certain functional domains ("Dialektabbau": dialect shift). Romaine (1986) presents a comparable analysis for languages-in-contact situations. Verhoeven & Boeschoten (1986) present evidence for the existence of a stage intermediate between acquisition and attrition, which they call "stagnation", i.e. a standstill in L1 acquisition due to simultaneous L2 acquisition.

Although the different types of language attrition research identified above exhibit many dissimilarities, it should be noted that there are also quite a few characteristics and research interests that they have in common (cf. Pan & Berko-Gleason 1986); to mention just a few:

- the influence of the competing language (variety) that either 'pushes out' (parts of) a language (variety), or 'fills in' the weak spots it starts to exhibit;

- the development of adequate assessment techniques able to cover a wide range of language proficiency levels;
- the discovery of more or less consistent patterns of attrition.

As Pan & Berko-Gleason (1986:204) noted:

> "It is perhaps too early to declare that the field of language skill attrition has become a unified subfield of linguistics, despite the fact that similar questions face all researchers studying language loss. It is clear, however, that linguistic theory, if it is to be complete, will ultimately have to account for the divergent phenomena observed when language skills are lost".

The phenomenon of *foreign*-language attrition deserves special attention because of the enormous investment of time and money in FL teaching, especially in a country like the Netherlands where virtually everyone learns at least *one* FL at school. Therefore, it is not only an academic question to investigate what happens to FL skills once formal instruction is over. Many FL learners will encounter - or may even seek - opportunities to use their FL skills, but others will, at least temporarily, not use them at all. The question is whether the latter group will gradually lose (some of) their skills over time and, if so, what is typically lost in such cases. This should not be seen as a mere evaluation of the long-term effectiveness of FL teaching, but also as a matter of interest to those who want to develop maintenance measures or 'refreshment' courses: they need to know what is lost before they can design maintenance or relearning courses (cf. Valdman 1982; Van Els & De Jong 1985). We will discuss the questions of what is lost and why, and how much is lost, in greater detail in the following two subsections (1.1 and 1.2). Separate sections are devoted to relearning (1.3), and to theories of forgetting (1.4).

Quite a different question is whether a shift in attention in terms of teaching content may - while teaching still continues - already result in the attrition of skills or knowledge not taught explicitly anymore. This problem is noted by Weis (1986) and Koster (1987). The latter concludes with respect to the perception of English sounds by Dutch university students of English that "without the benefit of language laboratory training or of practice in interpreting native speakers' speech, students tend to relapse in their ability to

perceive English sounds" (p. 82). This has, of course, important consequences for the organization of teaching programmes:

> "This suggests that the common practice at universities and teacher training colleges in Holland of giving foreign-language students intensive training at the beginning of their study and assuming that passing a proficiency test after one or two years is sufficient to guarantee a certain proficiency level at the time of their graduation, needs rethinking" (p. 136).

1.1. THE NATURE OF ATTRITION: WHAT IS LOST AND WHY?

Very little hypothesis testing has taken place on this point; we are in fact still in the stage of generating hypotheses, rather than testing them. Nevertheless, some attempts have been made at formulating relevant and testable hypotheses, usually on the basis of findings from research areas other than language attrition. Freed (1980:6) categorizes these attempts as follows:

> "(1) hypotheses based on regression theory: that is, a view of language loss as an unfolding or unraveling in reverse order of previously acquired forms;
> (2) hypotheses based on affective variables related to language learning and language maintenance;
> (3) hypotheses based on normative data of linguistic features controlled by proficient users of a language".

The *regression hypothesis* originates from Jakobson's (1941) monograph on the parallelism between diachronic language change, first-language acquisition, and language loss as a result of brain damage. Jakobson's actual data were relatively few, and all the 'evidence' available was in fact phonological. The most important attempt at testing the tenability of the regression hypothesis for aphasia is the collection of papers in Caramazza & Zurif (1978). Their general conclusion is that, except for segmental speech perception, it is untenable. This outcome is not surprising, however, in view of the fact that local brain damage produces a situation which - almost *a fortiori* - cannot be the 'mirror image' of child language acquisition: it results in *specific* deficits of parts of the language system rather than

a *global* deterioration of cognitive and linguistic skills, and it causes, in most cases, an *immediate* rather than a *gradual* deterioration (cf. De Bot & Weltens 1988).

In the context of L2/FL attrition - where the problems just mentioned do not seem to apply - there seems to be a much better chance of proving the existence of regression. The only studies that have - although only marginally - tried to do that are Cohen (1975), Hansen (1980), and Godsall-Myers (1981). In his multiple case study of the retention of Spanish as a FL in three American children over the summer recess of three months, Cohen did find some "examples to support the notion that some of the things that are learned last are also the first to be forgotten when the learners are removed from second [i.e. foreign] language contact for a period of time" (Cohen 1975:136). He adds, however, that not all data fit this pattern: he also found "new incorrect patterns", i.e. patterns that had not occurred at some earlier point in the learning phase, and "residual learning", i.e. "a reduction in certain problem areas. (...) some sort of unlearning of incorrect patterns during a respite" (p. 137).

Hansen's (1980) analysis of the learning, forgetting, and re-learning of Hindi-Urdu negation patterns by English-speaking children, showed that the attrition sequence was "a recapitulation in reverse of the acquisional sequence" (p. 169). Also Godsall-Myers (1981) claims to have found regressional patterns in the attrition of FL German in American students, but her claim seems to be based mainly on the fact that tests measuring simple phenomena - supposedly taught *and* learned early in the learning process - show less attrition than tests measuring more complex phenomena - supposedly learned much later. In other words, she assumes that difficulty level, order of presentation, and learning order are strictly parellel, and hence concludes that attrition increasing with difficulty level represents regression.

Attention may also be drawn to Ervin-Tripp (1974:118), who mentions the following anecdote:

> "By chance, we encountered two American children who were losing English after nine months living with their [French-speaking] Swiss mother and grandparents in Geneva. (...) They had regressed to a simpler sentence processing heuristic in which the cue from the function words and suffixes was inoperative, and the primary pattern, NVN=SVO, reappeared".

In all, the 'evidence' for regression in L2/FL attrition is rather scarce, and rather thin too. A promising way of looking at this process is a project proposed by Jordens et al. (1986): they are going to look at *one* phenomenon, German case marking, in L1 acquisition, FL learning, L1 attrition, and FL attrition. The German case marking system is an example *par excellence* on which to test the regression hypothesis, since it is acquired by native speakers in four different, quite discrete stages (no marking; nominative marking; nominative and oblique marking; nominative, dative and accusative marking).

Freed (1980), among others, notes that the regression hypothesis has occurred in a number of different variants (see also De Bot & Weltens 1988). One of these, 'best learned - last forgotten', concentrates on the *quality* of learning, rather than the *order* of learning. This particular interpretation of regression echoes the psychological 'levels of processing' theory (Craik & Lockhart 1972), which also links up retention with quality of learning (see e.g. Craik & Tulving 1975).

On the other hand, there is one instance of counter-evidence, namely in Brewer-Bomar (1981), who studied the attrition of L1 under the influence of acquiring an L2. Her conclusion was that quite the opposite of regression may occur:

> "[Although] interference was expected to first affect the grammatical categories last learned by the informants, almost the reverse was true. Some of the most basic syntactic patterns were the most interfered with, while semantically and grammatically more complicated models were not only left untouched in the L1, they were still being perfected".

In the second group of hypotheses, the *affective variable hypotheses*, the research findings on the role of attitudes and motivation in language *learning* (cf. Gardner & Lambert 1959, 1972), have been transposed to the area of language *attrition*; Gardner (1982:31-32) argues:

> "(...), *since attitudinal/motivational characteristics are related to the level of second language proficiency, they will relate to second language retention* (...). The model proposed argues that attitudinal variables involving integrativeness and the learning situation influence and maintain levels of motivation which, in turn, effect differences in second language proficiency (...). Attitudinal/motivational variables could

> also influence second language retention by orienting the individual to take every opportunity to maintain proficiency in the language".

It is exactly these results that Gardner et al. (1987:42) reported three years later in their study on L2 French attrition in Canadian Anglophones during the summer recess:

> "First, Language Attitudes are seen to 'cause' Motivation. (...) Second, Motivation is shown to be a causal factor in determining second language achievement and 'Use' of the language. (...) Motivation is shown to play a role in how much students attempt to use the language during the summer, and it is this 'Use' and the prior achievement that is responsible for individual differences in proficiency at the end of the summer".

More or less similar findings are reported by Gardner et al. (1985) for self-reported language attrition in a comparable population, and Snow et al. (1984) in their study of graduates of a Spanish immersion programme in the U.S.

Apart from these three recent studies, the references to the role of attitudes and motivation have been relatively anecdotal. Kennedy (1932:135), for example, noted that "intention to continue with the study of Latin is a very important factor in terms of the amount of initial knowledge retained over the summer vacation". Edwards (1976) did not find a (direct) influence of attitudes and motivation on retention, but - in line with Gardner et al. (1987) - he found that "those subjects with a higher language competence would tend to seek out more opportunities to use their skills" (p. 308).

The third group of hypotheses concerning the nature of attrition, the *linguistic feature hypotheses*, was formulated by Andersen (1982). Again, the basis for the hypotheses was found in other areas of research into language acquisition and use, viz. pidgin and creole studies, second language acquisition, language contact in bilingual communities, and language death. In general, Andersen's hypotheses centre around two factors: (1) (absence or presence of) contrast between L1 and L2/FL, and (2) frequency/markedness/functional load; in other words, the relation between elements in the two languages involved, and the relation between elements *within* the language system that is subject to attrition respectively.

As De Bot & Weltens (1988) note, the relation between these 'linguistic attributes' and the regression hypothesis discussed above is an ambivalent one. On the one hand, they may make parallel predictions: when, for example, an FL coursebook 'orders' its lexical items according to frequency, both 'last in - first out' and 'frequency' would predict that the retention of elements will be (inversely) related to their place in the coursebook. On the other hand, they may result in conflicting predictions, for example when two languages resemble each other more in the low-frequency than in the high-frequency domain: normally, these elements would be learned rather late in L2 because of their low frequency, and would, therefore, be early candidates for attrition under the regression hypothesis, whereas the linguistic feature hypotheses would predict a high degree of retention because of the similarity between the L1 and L2 elements.

Another source of interference is the fact that particular language elements may enjoy a special status in memory, e.g. closed-set items like the days of the week, the numbers of one to ten, and - Berko-Gleason (1982:21) adds - "songs and emotionally laden words like curses and body parts". Data from studies on L2 attrition in children by Berman & Olshtain (1983) and one of our own pilot studies (Aertssen et al. 1985) seem to support this idea: both studies observed severe lexical and grammatical attrition side by side with high retention of idiomatic phrases like *it's kinda hot* and conversational fillers like *let me see* and *for that matter*. Berman & Olshtain (1983:233) conclude:

> "(...) there are certain types of knowledge which are deeply entrenched through the original learning experience, where English [the L2] was acquired to the point of native-like proficiency in a naturalistic setting at school, with friends, in the neighbourhood, and often at home with siblings even though not with parents. Such knowledge seems very resistant to loss, especially among the older children".

An entirely different matter is whether attrition affects separate skills (reading, writing, listening and speaking) and/or discrete linguistic levels (phonology, lexicon, morpho-syntax) differently. Berko-Gleason (1982:21) notes:

> "(...) the traditional linguistic subsystems (...) may suffer differential loss in attrition, since they are learned separately. It is also generally accepted in the child language world that in all cases, comprehension

> precedes production (of systematic aspects of the language, not individual instances)".

Studies that have measured proficiency in different skills and/or on different linguistic levels have in many cases reported skill-related or level-related differences. Edwards (1976) found an 8% gain in reading after 12 months of non-use, but a 13% loss in speaking. Geoghegan (1950) found a complex interaction between target language and skill/level: gain in French vocabulary; loss in translation, vocabulary, and grammar for Latin, and loss in translation and vocabulary for Spanish. Skill-related differences were also reported by Scherer (1957), Pratella (1969), Smythe et al. (1973) in their first of two experiments, Godsall-Myers (1981), and Bahrick (1984). Finally, Edwards (1977:58) notes a "trend towards decline in speaking ability" while there is not "any appreciable loss in reading, writing and listening".

In one of our own pilot studies on FL vocabulary loss (Messelink & Verkuylen 1984, summarized in Schumans et al. 1985) we found that recall is affected more heavily than recognition; in fact, this is the only more or less consistent finding: productive skills seem to be more vulnerable than receptive skills (but cf. Clark & Hecht 1983).

1.2. THE RATE OF ATTRITION OVER TIME

There are several reports witnessing that attrition can be particularly fast in (young) children; accounts of such rapid (L2) loss can be found in, for example, Hansen (1980) and Olshtain (1986). There is even an (anecdotal) report on the rapid attrition of L1 skills in a 4-year-old child; Burling (1978:70-71) writes:

> "there was still no doubt that Garo was his first language (...) but within six months of our departure, he was even having trouble with the simplest Garo words, such as those for the body parts, which he had known so intimately".

The situation with FL attrition is, of course, very much different, in the sense that FL learners usually are adolescents, if not adults.

Most of the 'older studies' on FL attrition have only compared two points in time: pre- and post-summer vacation scores. Their findings are

ambiguous: they run from substantial gains in some areas (among others, Cohen 1975 and Edwards 1976) to losses up to 34% over the same period of time, i.e. three months (Kennedy 1932).

In order to gain an insight into the progression of attrition over time - enabling us to draw a forgetting curve, or its complement, a retention curve - we need, of course, at least three points of measurement. The oldest observations on this point date back to the nineteenth century, to the 'traditional forgetting curve' by Ebbinghaus (1885; cf. MEERLING 1981:13-17), which essentially predicts two things:

(1) attrition sets in rather quickly, but attrition rates decline exponentially in subsequent periods;
(2) attrition is proportionate to the original proficiency level; in other words, the *percentage* of knowledge lost in a given period of time is independent of the original level.

The first prediction was confirmed by Kennedy (1932), Flaugher & Spencer (1967), Godsall-Myers (1981), and Bahrick (1984), who all found heavier attrition in initial periods of non-use than in subsequent ones. On the other hand, there are a number of studies that have observed the opposite: attrition does not set in immediately; there is an initial 'plateau' before retention declines. Edwards (1977) found no loss after 6 months, but significant loss after 12 and 18 months; Messelink & Verkuylen (1984; see also Schumans et al. 1985) found no loss after 1 year, but a loss of 15% after 2 years of non-use, and Snow et al. (1984) and Grendel & Poppe (1986; also reported in Weltens et al. 1986b) did not find (significant) losses after 2 years of non-use, but they did after 4 years.

In other words, there are basically two different patterns that have been observed. They have been visualized in Figures 1.2a and 1.2b.

It should be noted, however, that the two are not mutually exclusive. It might well be that the pattern of Figure 1.2a will come after the initial 'plateau' with which Figure 1.2b starts, but that no study in which the plateau occurred covered a time-span long enough to reveal this pattern.

It is quite remarkable that the four studies that reported the initial plateau all employed subjects of high proficiency compared to the subjects in most other studies: English-dominant bilinguals in Canada (Edwards 1977), graduates of a Spanish immersion programme in the U.S. (Snow et al.

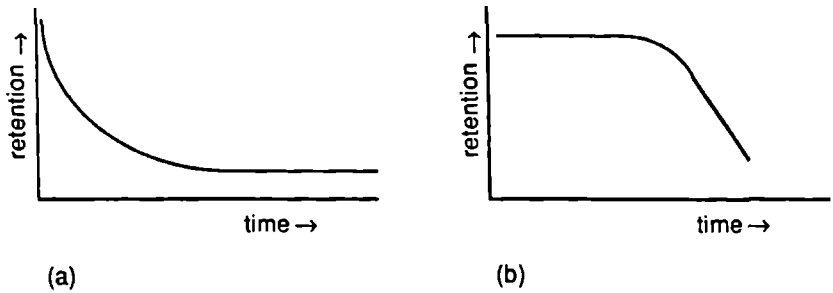

Fig. 1.2: Two possible retention curves.

1984), and Dutch FL students with four and six years of relatively high-standard FL training respectively (Messelink & Verkuylen 1984; Grendel & Poppe 1986). In other words, it might be that the Ebbinghaus curve does generally fit FL attrition data, except when we are dealing with relatively high levels of proficiency: in those cases the curve will be preceded by a plateau, i.e. a period of time in which the proficiency does not degrade - or does so only non-significantly. This option seems quite likely in view of Neisser's (1984:33) reaction to Bahrick's (1984) study: it might be that there is a "critical threshold during learning" beyond which isolated responses, or 'facts', become part of "mental representations of complex information structures" with the result that they - at least temporarily - "become immune to interference or decay"; a similar conjecture is made by Pan & Berko-Gleason (1986:204): "Is there a critical mass of language that, once acquired, makes loss unlikely?".

There are, in fact, indications that proficiency increases somewhat during the first months after course completion. We have already mentioned Cohen's (1975) "residual learning", and the gain in overall proficiency after 3 months of non-use observed by Smythe et al. (1973), but we have come across the same phenomenon ourselves in one of our pilot studies on the retention of FL skills after 2 years of non-use (reported in Weltens & Van Els 1986). Our subjects seemed to have improved significantly (at the .05 level) in general proficiency in French as measured by a multiple-choice cloze test (cf. section 2.5.1). The explanations suggested for this phenomenon in Weltens & Van Els (1986) centre around the idea that the subjects may have matured in the meantime - cognitively in general, or just in terms

of reading experience, test-wiseness, etc. - so that their score on the test in question could become higher, although other tests suggested that their lower-level skills had suffered somewhat. In other words, general proficiency tests may not be the most adequate to discover (early) signs of attrition, because they generally allow the use of all kinds of compensatory strategies to make up for any lexical or grammatical deficits that have arisen. In addition, as Jaspaert & Kroon (1987) have argued, general proficiency tests are relatively susceptible to 'testpertise' effects: variation in test scores at least partly reflects variation in testpertise, rather than just variation in actual proficiency.

Both Scherer (1957) and Cohen (1975) suggest that some 'post-course processing' may occur; the former remarks:

> "There is some reason to believe that the five-hour course does not afford the busy student enough time for full digestion of the material. His mental metabolism continues to work after the course is over and finally catches up at some time during the summer" (Scherer 1957:257).

On the other hand, Smythe et al. (1973:405) suggest that "students may be 'fresher' and more motivated to do well on these tests after the summer vacation". A second suggestion they make concerns what psychologists call 'spontaneous recovery of information stored in memory' (see, for example, Loftus & Loftus 1976:77), but they add that it remains unclear why it should only apply to listening comprehension, but not to reading comprehension; and why it should only occur after 3 months, but not after 8.

It should be noted that 'residual learning' and 'post-course processing' refer to a (subconscious) continuation of the learning process, whereas 'spontaneous recovery' refers to the restoration of the link between a stimulus and a response that was learned at some point, but extinguished afterwards.

The second problem we noted in connection with forgetting curves is *the role of the original proficiency level*, that is the proficiency level attained by the end of formal instruction. Theoretically, the three most obvious relationships possible are the following: a positive relationship ("The more you know, the more you lose"), a neutral one ("You lose a fixed 'amount' irrespective of your total knowledge"), or a negative one ("The more you know,

the less you lose") It is important to note that matters vary according to whether attrition is defined in *absolute* or in *relative* terms, i.e. in raw scores or in percentages of the original level. This is represented in Figure 1.3, where the three relationships have been visualised in absolute terms, and translated into relative terms. (For the sake of clarity, we have drawn the figures as if attrition proceeds in a linear fashion.)

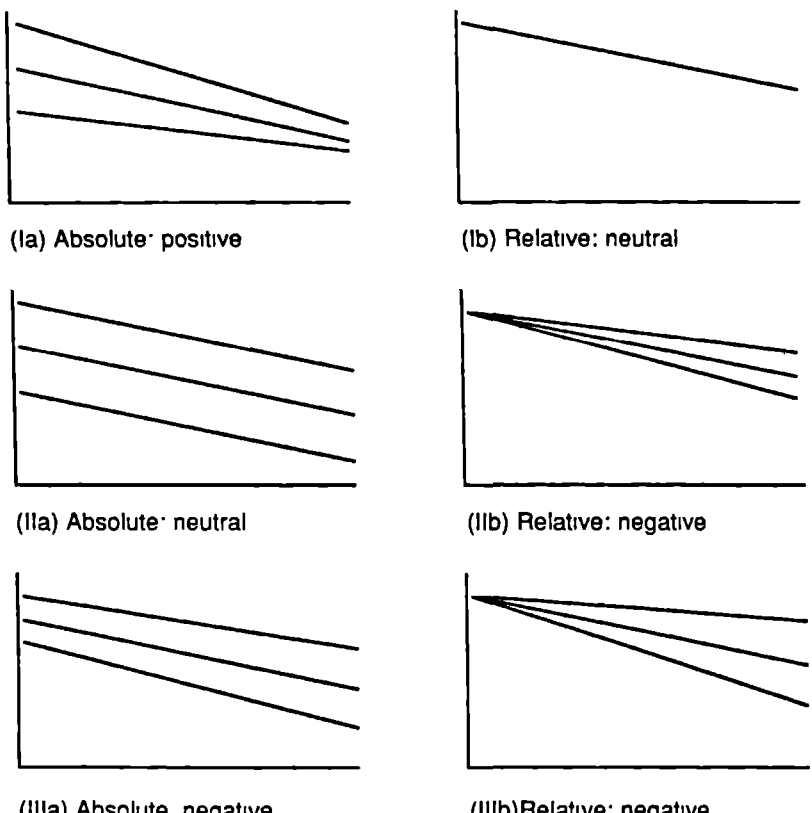

Fig 1 3 Possible relationships between original proficiency level and attrition

The first relationship (Ia) is one in which higher proficiency levels lose more, in relative terms (Ib) this means a neutral relationship: the *proportion* of knowledge lost over a given period of time is equal for all proficiency levels (with the result that all three lines do in fact coincide). Along this line Kennedy (1932:146) concluded that "initial achievement is the significant

factor in retention", and Scherer (1957:276) summarized this possibility as follows: "Perhaps this is the case because good students had more to forget". Indeed, this remark echoes the well-known metaphor of the wine barrel and the 'leaking-parameter': the more wine there is in the barrel, the harder it leaks. (Note that this metaphor not only 'explains' that high proficiency subjects lose more in absolute terms, but also that attrition rates decline in subsequent periods: as the content of the barrel becomes less, it gradually leaks more slowly.)

In the second relationship (IIa), the amount of attrition is independent of proficiency level: subjects lose a fixed amount of knowledge in a given time interval. Consequently, high-proficiency subjects lose relatively less; in other words, there is a negative relationship between proficiency level and attrition rate (IIb). This finding was obtained by Smythe et al. (1973:405), who reported the "absence of any differential forgetting rate as a function of grade level" - but apparantly they mean "amount" instead of "rate". Godsall-Myers (1981:59) also found attrition "in inverse proportion to" proficiency level: the absolute attrition percentages do not differ very much, but relatively to the original proficiency level, the pretest score, they do: the attrition rates tend to increase with lower levels of proficiency. Along the same line, Bahrick (1984:116) concluded that:

> "the total *amount* of content to be forgotten during the five years following training is relatively constant for individuals at different levels of training, but this amount becomes a progressively smaller portion of total knowledge with higher levels of training".

An important consequence of this position would be that "very low levels of proficiency resulting from short and miniscule FL programmes disappear rapidly and completely after a relatively short period of non-use" (Van Els 1985).

The third relationship, then, a negative one both in absolute (IIIa) and in relative (IIIb) terms, was - as yet - only found by Pratella (1969) and Robison (1985). Pratella (1969) compared FLES (Foreign Languages in the Elementary School) students, who had begun studying Spanish in grade 5, with non-FLES students, who had begun in grades 9, 10, or 11. Both groups showed significant attrition over a period of 3 months, but attrition was much heavier in the lower-proficiency group, the non-FLES students.

Robison (1985) investigated a comparable population over the same

time interval, and also found that - at least for syntax and vocabulary - retention was higher for more advanced students.

An additional problem in determining how much is lost over a given period of time is the fact that many studies covered a time period of only three months - as will have become clear from the discussion above. As we have stipulated elsewhere (Weltens 1987), it may well be that we have to look at much longer periods of non-use in order to arrive at a clear understanding of the processes involved in language attrition, the relationship between original proficiency level and attrition rate being just one of them. Also, some studies only hint at the role of post-course exposure (e.g. Edwards 1976, 1977), although this is a central factor in other people's views. Gardner et al. (1987:45) summarize both points quite clearly and concisely: "Future studies should attempt to extend the length of the incubation period [i.e the period of non-use] while ensuring that students do not participate in formal language study". As we hope to show in chapter 2, this is exactly what we did in our investigation.

1.3. REMEDYING LANGUAGE ATTRITION: RELEARNING

Although it seems to be generally understood that one of the main applications of language attrition research lies in the area of relearning, the attention paid to this phenomenon has been surprisingly small. One of the very few studies that pay explicit attention to it is Kennedy (1932). He not only looked at the attrition of Latin syntax during 3 and 12 months following training, but also at the effect of resuming training after 3 months: he found that one month of renewed study was "more than equal to the task of returning this group to its initial level of achievement" (p. 141).

Weis (1986) investigated the reactivation of FL vocabulary through the administration of a test, and a one-hour discussion of the test results. His findings suggested that this activity alone increased the (productive) availability of the particular vocabulary items included in the reactivation activity by over 100%: after 12 months - in which learning continued, but no particular attention was paid to the set of vocabulary items under investigation - only 24% was available productively; one month after the reactivation activity this had increased to 49% of the original set.

The studies on L2 attrition by Allendorff (1980) and Hansen (1980) also explicitly included relearning in their object: they both found that relearning

was much faster than first-time learning. It is problematic, however, that both studies employed very young children, so that the faster relearning process may - at least partly - be due to maturation on the part of the children, as Hansen (1980:157) rightly points out herself.

An important psychological claim in this respect is that forgetting does not consist in *loss* of information from memory, but in *inaccessibility* of information. In other words, nothing that enters long-term memory ever seems to be removed from it again (Gregg 1975; Loftus & Loftus 1976; Thomassen & Kempen 1976). Hence the tendency to prefer the term 'attrition' over 'loss'.

The difference between loss and inaccessibility of information is of particular relevance for the potential application of attrition research in refreshment courses. If loss does *not* occur, refreshment means *re*-learning, i.e. a process that will definitely be faster than (and different from?) first-time learning. This may also have consequences for *which* material should be presented in such courses. In a research project recently started, De Bot & Schreuder (1987) are investigating the effect of a reactivation training on lexical skills in FL French. On the basis of Cohen (1986), they distinguish a number of aspects of vocabulary knowledge, which they first test after 18 months of non-use, then retrain inasfar as they have been lost, and subsequently retest. Their main research question is whether reactivation will only have an effect on the material actually trained, or also on related material (De Bot & Schreuder 1987:3).

Two interesting suggestions for actual relearning techniques come from Clark & Jorden's (1984) subjects. The first suggestion - which, surely, echoes the experience or at least the intuition of a good many FL learners - concerns a short return to the L2 environment, about which a number of subjects remarked:

> "that relatively short, immersion exposure to the language upon their return to Japan allowed them to regain very quickly a substantial amount of their original ability, at least for the kinds of language skills at issue in daily interaction contexts" (Clark & Jorden 1984:55).

As Campbell & Schumann (1981:83-84) noted, this phenomenon is well-known in "the intuition of laymen", but it has not "been documented with any degree of seriousness". None the less, it supports the psychologists' view of memory: information is not lost, but becomes inaccessible, and can

therefore be reactivated relatively easily. (Campbell & Schumann 1981 suggest that hypnotism could be a tool for gauging the amount of information actually contained in a subject's memory, but as yet there is only anecdotal evidence that this is the case indeed.)

The second suggestion in this respect concerns the materials to be used in a refreshment course, namely:

> "the observation by at least one student that the review of familiar textbooks would provide for more rapid and more thorough re-acquisition than would a corresponding amount of work with new materials" (Clark & Jorden 1984:55).

Both these suggestions fit in with a theory of forgetting that interprets forgetting as retrieval failure: retrieval of information is facilitated when the information is 'looked up' in the context in which it was originally acquired. This context in itself is a powerful 'retrieval cue', resulting in more - or better, or faster - retrieval (cf. Loftus & Loftus 1976:82-84). We will return to the *retrieval-failure theory* in more general terms in the next section (1.4).

1.4. THEORIES OF FORGETTING

Obviously, it would be beyond the scope of the present book to treat the entire psychological literature on forgetting in detail here. Our discussion will be limited to some of the main characteristics of the major theories on forgetting, insofar as they are potentially relevant for our research.

As we have argued elsewhere (Weltens 1987:22), there is no reason to assume that non-use of a foreign language - i.e. the total absence of activation of this knowledge - in itself should result in attrition, unless one wishes to accept the so-called *decay theory* or *trace-fading theory*, which was formulated at the beginning of this century, but has been shown to have only very limited explanatory value, if any at all (see e.g. Kolk 1974:11-12; Thomassen & Kempen 1976:375). The theory claims that memory traces decay if they are no longer used: "disuse leads to something like physiological atrophy of the traces" (Kolk o.c.:11). Thomassen & Kempen (o.c.:375) point out that decay is untenable as the *only* explanation for forgetting, "because it has been shown in many ways that cognitive activity influences the forgetting process both retro- and proactively" (our

translation). The main criticism against the decay theory is that it has been shown that a state of total cognitive inactiveness is rare, if not non-existent, for human beings; hence the idea that the explanation for forgetting should be sought in the intervening activities, rather than in the non-use of the material that had been forgotten. This led to "one of the most widely held explanations of forgetting", the *interference theory*: "people forget an event because something else they have learned prevents the event from being remembered" (Loftus & Loftus 1976:74). A distinction is usually made between situations in which previously learned material interferes with newly learned material, *pro-active interference*, and those in which newly learned material interferes with previously learned material, *retro-active interference*. In the context of FL learning, a further distinction is very common, namely that between positive transfer from L1 to L2 (*facilitation*) and negative transfer from L1 to L2 (*interference*) (see e.g. Van Els et al. 1984:49-52).

The second major theory of long-term forgetting is the *retrieval-failure theory*,

> "which is more in line with the information-processing approach, [and] views forgetting not in terms of unlearning or competition of conditioned responses, but as a failure to retrieve some desired information. (...) forgetting is much like being unable to find something that we have misplaced somewhere. Forgetting occurs because the information we seek is temporarily inaccessible; if only we had the right retrieval cue, the information we seek could be successfully retrieved" (Loftus & Loftus 1976:78).

The experimental paradigms used in research dealing with one of these - and other - theories usually differ substantially. Therefore, it is not surprising that none of them has yet qualified as *the* theory of forgetting. Most accounts of forgetting conclude by noting that either theory seems to contribute somewhat to our understanding of human memory. Even a handbook such as Loftus & Loftus (1976), which devotes a relatively large number of pages to long-term forgetting, brings up only one single experiment in which the two major theories were tested concurrently (pp. 83-84): Richard Shiffrin carried out an experiment in 1970 in which he had subjects recall word lists of either 5 or 20 items. Between presentation of the list to

be recalled and the actual recall test, he presented intervening lists of words which also contained either 5 or 20 items. The interference theory would predict that the length of the intervening list would affect recall, whereas the retrieval-failure theory would predict recall to be dependent only on the length of the list to be recalled. Shiffrin's results were quite clear: "Recall is determined by the length of the list to be recalled and is independent of the length of the intervening list, just as is predicted by a retrieval failure theory" (p. 83). But although this experiment clearly supported the retrieval-failure theory, the authors are hesitant to reject the interference theory "because the theory has been designed to cover different situations" (p. 84).

An additional problem is that both theories - like most psycholinguistic theories - were based on carefully controlled 'verbal-learning' laboratory experiments, in most cases employing listwise learned nonsense words (see e.g. Postman 1971), rather than on evaluations of language skills (cf. Levelt 1978; Van Els et al. 1984). One may well wonder, with for example Kolk (1974:13), whether these theories have anything to do with real-life forgetting, let alone with the forgetting of such a complex phenomenon as language: Cofer (1984:30) notes that studies of long-term retention usually concerned "retention for materials that (...) contained few or no principles or systematic features (...) [whereas] any natural language contains structural and systematic features that once acquired serve as the foundation for the use of language".

Quite indicative of the inapplicability of the psychological treatment of forgetting to language attrition is the fact that there is - to our knowledge - only one passing reference to these theories in the literature on foreign-language attrition, namely the conclusion drawn by Smythe et al. (1973:405) that "the absence of any differential forgetting rate as a function of grade level is unexpected in respect to the interference position. That is, students in the higher grades, having shown a higher degree of initial learning, should have been less affected by interference". They refer, in fact, to just about the only aspect of our investigation that theories of memory would make any predictions about: the influence of the original proficiency level (cf. section 1.2 above). The other questions addressed here - are different skills and/or different linguistic levels affected to the same degree and in the same way?; are there particular 'subsets of elements' which are less vulnerable than others? - simply do not figure in either theory.

2. THE RESEARCH PLAN

In this chapter we will present our research plan. We will describe the aim of the investigation (2.1), the choice of the target language (2.2), the design (2.3), and the skills and linguistic levels tested (2.4). Then we will describe our instruments (2.5) and our informants (2.6). Finally, we will outline the procedure followed in the test sessions (2.7).

2.1. THE AIM OF THE INVESTIGATION

The situation where people learn a language and subsequently do not use it for quite some time, is in fact inherent to the Dutch school system. Most secondary school students initially learn three (modern) foreign languages as obligatory subjects, viz. English, French and German. For the exams they choose a total of six or seven exam subjects - depending on the particular school type - with the result that one or two of the foreign languages may be dropped two years before the school-leaving exams.

In the highest type of secondary school, the so-called *VWO* - the one that prepares mainly for university, which is also the one that we are dealing with in the present study - students choose seven exam subjects after four years of secondary school, i.e. at the end of what we will henceforth refer to as "secondary-4" (SEC-4). English is chosen as an exam subject by 98% of the students; German and French are chosen by about 50-55% and 35-40% of the students only (Ginjaar-Maas 1985:7; Oud-de Glas 1985:31). This means that many students drop German and/or French two years before their final exams, and - as a consequence - enter university or professional training, where they may again need these foreign languages, after a period of two years in which they have had hardly any exposure to them. (Incidentally, we will refer to students who continue to study the FL as "choosers", and to those who drop it after SEC-4 as "non-choosers".)

The main questions in this research are:

(1) whether, and if so, to what degree, foreign-language proficiency 'attrites' during longer periods of non-use;
(2) whether different skills and different linguistic levels are affected to the same degree, and whether they exhibit the same pattern of

development over time;
(3) whether there is a relation between the proficiency level attained and the rate of attrition; and
(4) to what degree the relation between (rules and elements in) the L1 and the FL plays a role in the attrition processes.

We feel that these questions are relevant in terms of foreign-language teaching policy: one of the main objectives of foreign-language teaching in secondary schools is to endow students with skills which they will not be actually using until *after* they have finished school, be it in higher education, be it professionally or otherwise. Many people know from experience, however, that foreign-language skills may 'attrite' relatively fast. In the Netherlands, this belief is particularly persistent for French, as is confirmed by the results reported in Claessen et al. (1978:267): former secondary school students overwhelmingly report that their school-learned French has deteriorated since they left school, irrespective of the type and level of secondary school they attended.

In a U.S. context, similar problems are reported with respect to "student placement in French": Hagiwara (1983:27) notes that students enrolling for French courses at the university often have a number of "intervening years" between their FL training in high school and the moment of enrolment, with rather serious consequences:

> "(...) student scores on the placement test, in every case examined, declined steadily as the number of interrupted years increased, and (...) two or more years of interruption resulted in a loss of more than one quarter of the original proficiency. (...) other surveys have shown consistently the same pattern: the longer the intervening period, the lower the placement".

However, we do not know *how fast* this process takes place, *what* is typically lost, whether attrition follows some 'universal' pattern, and whether this attrition is definite - to mention just a few of the most obvious questions that may crop up (cf. also Lambert & Moore 1986). Therefore, research into foreign-language attrition may - directly or indirectly - contribute to an evaluation of the effectiveness of foreign-language teaching in the present educational system, in particular with respect to the situation in which students have the opportunity to drop a FL after four years of training (see

also section 2.3 below).

2.2. THE TARGET LANGUAGE: FRENCH

As noted in our introductory chapter, the control of intermediate exposure is rather problematic. One way around this problem is the *elimination* of this variable, in other words by ensuring that the subjects are not exposed to any relevant input once they have left formal training. In our research, the choice of the FL to be investigated was partly determined by this consideration. Theoretically, we had three languages to choose from: English, German, and French. As already mentioned in the previous section, there are virtually no non-choosers of English, and, in addition, English is so pervasive in the Dutch media that "non-use" is effectively impossible. Therefore, we had to choose either German or French. Now, *total* lack of exposure is much less likely for German than for French, because of the fact that almost everybody in the Netherlands can watch German television, or listen to German radio. This is, of course, particularly true of the town where we carried out our investigation, Nijmegen, which is just a few miles off the German border. Therefore, we decided to investigate the attrition of *French* as a foreign language: once students have left formal training, they will most likely have hardly any contact with French in subsequent years.

2.3. THE DESIGN OF THE INVESTIGATION

As was made clear above, investigating a period of two years more or less 'naturally' follows from the Dutch educational system. In line with our criticism that many previous studies looked at only *one* period of non-use, we wanted to include at least two in our own project. Given the nature of the funding of this research, however, the time interval that could be covered longitudinally was limited to two years. Therefore, we decided to stick to the period of two years, but to study one such interval longitudinally, and a subsequent one cross-sectionally, so that a total time lapse of four years could be covered.

The levels of proficiency to be investigated also more or less followed from the Dutch educational system: since there are basically two 'types' of students - those with four years of FL training and those with six - these seemed to be the most obvious levels to be included in our project. The difference between the two happens to be exactly 50% of training: for French

the number of lessons per week is 3 on average in each grade; therefore, students who choose French receive 50% extra training in SEC-5 and SEC-6, on top of what all students receive in grades 1 through 4 (see e.g. Van Els 1981). Although, for example, Seliger (1985) heavily opposes to a definition of proficiency in terms of semesters or years of training, and argues for an operationalisation based on "actual measured language knowledge of some kind" (p.16), we are quite confident that this difference in training results in significantly different levels of proficiency. This was confirmed in our pilot studies, which showed that - at least for the lexical and grammatical levels - the increase in knowledge in the upper two grades is considerable (cf. Schumans et al. 1985, and Weltens et al. 1986b respectively). We should add here that Seliger (1985) is primarily referring to what might be labelled 'retrospective' studies of language attrition, in which proficiency at the end of training - i.e. the baseline - is defined solely in terms of years or semesters of training. In our case, the baseline was established by actually measuring our subjects' proficiency at that point in time.

When we combine the three previous arguments, i.e. an interval of two years, two such intervals, and two levels of training, we arrive at the design shown in Table 2.1.

Table 2.1: Design of the research project (The arrows indicate longitudinal measurements).

		Years of non-use				
		0		2		4
Years of training	6	A		B	-->	C
	4	D	-->	E		F

Since the moment of the final exam for secondary school students (at the end of SEC-6) is the moment when they spread all over the country to attend all kinds of further training, we thought it would be most practical to include this moment in the cross-sectional comparisons; in other words, the developments between groups A and B, and between groups E and F (see Table 2.1) were measured cross-sectionally; the developments between B

and C, and D and E were investigated longitudinally, i.e. the subjects in groups B and D were tested in April/May 1985, the first point of measurement, and retested at the second point of measurement, in April/May 1987, when they had become groups C and E respectively.

We are aware of the fact that there is a continual discussion of the pros and cons of longitudinal as opposed to cross-sectional designs (see e.g. Campbell & Stanley 1966; Baltes et al. 1977; Ginsberg 1986, and Jaspaert et al. 1986; the latter two deal with this problem in the specific context of language attrition research), but we will refrain from discussing that at length here. We are also aware of the fact that a combination of the two in a so-called *sequential design* (Schaie 1965) is in itself a recommendable design, but we rather combined them into what Campbell & Stanley (1966:57) have labelled a "patched-up design":

> "(...) a strategy for field research in which one starts out with an inadequate design and then adds specific features to control for one or another of the sources of invalidity. The result is often an inelegant accumulation of precautionary checks, which lacks the intrinsic symmetry of 'true' experimental designs, but nonetheless approaches experimentation. (...) characteristic of such designs is that the effect of *X* [the treatment] is demonstrated in several different manners".

In our case the threat came from the 'history' factor: could the different cohorts involved in the comparisons be compared? To be more specific: would it have made any difference had we observed all six groups in a purely cross-sectional design? In order to answer this question, a check for possible cohort effects was built in by testing another sample of subjects belonging to groups C and E - the ones that would be measured longitudinally - at the *first* time of measurement as well. These two control groups only took the multiple-choice cloze test, the test we used for measuring general proficiency in French. The combination of longitudinal and cross-sectional data will be evaluated in sections 3.1 and 4.1.

Another problem that is generally considered of central importance is the establishment of the baseline, i.e. the point of reference for determining what has been lost compared to an earlier point in time. Jaspaert et al. (1986) discuss this problem at length in relation to L1 attrition. They note that Andersen's (1982) notion of the *LC*, the "linguistically-competent

speaker", is inadequate in many research situations; a similar argument is presented by Seliger (1985). The LC is certainly inadequate for FL attrition, although it is occasionally even used there as well, for example in Weis' (1986) study, which implicitly assumes that mastery of a particular set of vocabulary items is 100% after a certain amount of training. How untrue this assumption may be, can be seen in, for example, Guntenaar & Willemsen's (1987) investigation into foreign-language attrition, in which they employed our written lexical test (cf. section 2.5.5 and Appendix 7). The French teacher of their subjects estimated that, at the end of SEC-4, they would know at least 50% of the low-frequency non-cognates tested (cf. Appendix 7); in reality, the choosers knew, on average, 18% and the non-choosers only 10% of these words. Similarly, the French teachers of our own subjects claimed they taught all of these words at least up to receptive mastery (cf. section 2.5.5), but in practice the choosers knew only 43% and the non-choosers 35% on average.

In our design the baseline is established by measuring subjects right at the end of the training period, i.e. at the end of grades 4 and 6. In this way, we feel, we establish a valid baseline to set off attrition against.

2.4. THE SKILLS AND LINGUISTIC LEVELS TESTED

The first choice we made was to limit the project to the receptive skills. There were a number of reasons for doing so. Firstly, there is some evidence to assume that if attrition can be shown to have affected the receptive skills, it will most probably have affected the productive skills (see e.g. Snow et al. 1984, and Schumans et al. 1985). Secondly, the receptive skills can be tested in a relatively easy way; measuring speaking and writing proficiency would entail all kinds of validity and measurement problems, and require laborious analysis of speech and text products - in case one would opt for a 'free', unguided type of test (cf. Nienhuis 1977; Yorozuya & Oller 1980; Van Weeren 1982). Thirdly, there are standardized tests available for measuring listening and reading proficiency, developed by the Dutch National Institute for Educational Measurement (CITO; cf. sections 2.5.2 and 2.5.3). Finally, although the official teaching objectives require an equal emphasis on all four skills, the receptive skills receive much more attention as far as the teaching of French as a FL in the Netherlands is concerned (see e.g. Van Els & Radstake 1987) - and rightly so, many would add, because they are the skills that people might want to use later on in

their professional careers. Moreover, some authors have argued that functional language skills develop best if teaching starts with receptive skills (see e.g. Schouten-van Parreren 1983).

We decided to test receptive proficiency in French on three different levels:

(1) general proficiency - as measured by e.g. cloze and dictation tests;
(2) listening comprehension and reading comprehension; and
(3) (within each of these two modes) the mastery of certain phonological, lexical and morpho-syntactic elements or rules.

Since there is some evidence for the existence of "some overall language proficiency" (Van Els et al. 1984:326), which is usually measured by means of cloze tests (Oller 1973; Wijnstra 1977) or dictation tests, we decided to include a test aimed at measuring this general proficiency in our project, side by side with tests measuring listening and reading comprehension. Although we share the scepticism about its validity expressed by a growing number of authors (cf. Lapkin & Swain 1977; Klein-Braley 1985; Lee 1985, Markham 1987), we opted for a cloze test, and since the entire project is concerned with the receptive skills, we decided to use a *multiple-choice* version, so as to avoid that subjects would actually have to *produce* French words.

A multiple-choice cloze test is also used in the (national) university entrance exam of Dutch for foreign students in the Netherlands. The reliability of the 100-item test is very high (over .90), but the population taking the exam is of course very heterogeneous compared to our subject sample (cf. 2.5 below). In this context, the MC cloze test consistently correlates reasonably well with the listening, speaking, reading and writing tests used, viz. in the range of .61 to .83 (Janssen-van Dieten & Raymakers-Volaart 1986). Incidentally, some people have been impressed by this to such a degree, that it has led them to argue that one might just as well do without the other tests, and use just the MC cloze test as the exam (Jochems & Montens 1987).

For level (3) we had to develop our own tests, six in total. In line with Andersen's (1982) *linguistic feature hypotheses* (cf. section 1.1 above), we decided to include the 'contrast' factor, i.e. the factor 'relation between L1 and FL', on all three linguistic levels. On the phonological level this meant testing the discrimination of native-like phonemes versus the discrimination

of FL-specific ones; on the lexical level this was operationalised as testing cognates versus non-cognates, and on the grammatical level as testing native-like grammar rules versus FL-specific ones (see sections 2.5.1 through 2.5.7 for a further elaboration of the test selection and construction respectively).

In the lexical tests we included a second "linguistic attribute", namely frequency. This was done for a number of reasons. Firstly, because there is a - for our purposes - exhaustive frequency count of modern French (Savard & Richards 1970). Secondly, because the frequency of lexical elements has been shown to play a significant role in FL learning (see, for example, Kerkman 1982, 1984, in prep.) and FL attrition (Messelink & Verkuylen 1984; Schumans et al. 1985; Verkaik & Van der Wijst 1986). Of course, we would have wanted to include similar factors on the phonological and morpho-syntactic levels, but there we did not have such a frequency count at our disposal, or a reliable way of defining degrees of 'functional load' of grammatical rules and phenomena. Therefore, we incorporated only one item factor in the tests for the phonological and the morpho-syntactic levels, namely 'contrast', and two in those for the lexical level, namely 'contrast' and 'frequency'.

To sum up, then, we used the following tests:

(1) a multiple-choice cloze test (cf. 2.5.1);
(2) a listening and a reading comprehension test (cf. 2.5.2 and 2.5.3 respectively);
(3) two phonology tests: a phoneme-discrimination test and a variant of the rhyme test (cf. 2.5.4); two vocabulary-translation tests (cf. 2.5.5); and two multiple-choice blank-filling tests for morpho-syntax (cf. 2.5.6).

2.5. THE INSTRUMENTS

As explained in the previous sections, six tests had to be developed in the course of the project. Some of the construction activities were carried out in the form of student projects. In the sequel, we will be referring to these projects as 'pilot studies'. They were the following. Van Agt & Wessels (1984) report on some preliminary work for the phonological tests. Messelink & Verkuylen (1984) describe the first pilot study that ultimately led to our lexical tests; the selection of the target words in these tests was largely

based on a subsequent experiment reported in Verkaik & Van der Wijst (1986). Finally, the development of the morpho-syntactic tests was based on the work reported in Grendel et al. (1985) and Grendel & Poppe (1986).

A general problem for our test construction - i.e. for the construction of the phonological, lexical, and morpho-syntactic tests - was test reliability. There were a number of things that *a priori* threatened the level of test reliability of our pilot tests:

(1) all these tests were pilot-tested with only 18 subjects;
(2) the group of subjects in the pilot test was relatively homogeneous compared to the total sample that would take part in the actual investigation;
(3) the tests by definition contained items that would probably be answered correctly by all subjects; such items do not contribute to the reliability of a test, because they do not contribute to its variance;
(4) the tests by definition did not consist of a unidimensional scale of items of equal difficulty, but of discrete subsets of items of considerably different levels of difficulty.

The third and fourth considerations do, in fact, echo a very fundamental problem: classical measures of test reliability were developed for norm-referenced tests, and not for criterion-referenced tests (see e.g. Popham 1978:89-111), such as the ones we are concerned with here. Van Els et al. (1984: 315-316) conclude on this point:

> "(...) the statistical framework for criterion-referenced testing is still being developed and has not yet reached the degree of maturity that classical testing theory for norm-referenced measurement has attained. One conclusion which can be drawn in any case is that the better it has been described in advance what mastery of an item means, the better the test scores can be interpreted".

One way out for the problem that these tests are not unidimensional, but consist of predefined subsets of items - leaving aside for a moment the question whether one should compute reliability coefficients for these tests at all (cf. Schils & Van der Poel, forthcoming) - is given in, for example, Stanley's (1971) extensive discussion of reliability analyses. He argues:

> "Where several items refer to the *same* unit (e.g. reading passage, graph, or table), the group of items in this unit, rather than the individual items should be the basis for split-halving or other internal consistency determination" (Stanley 1971: 409).

The problem then is, of course, that our tests would consist of very small numbers of 'items', namely two or four (see below); no one would develop a four-item test and expect it to be reliable. This problem was alleviated somewhat by running a reliability analysis across the *two* tests, i.e. by computing the composite scores for the subsets of items in either test (cf. section 2.4 above), and treating these scores as if they were items. The result was, then, one indication of the test reliability for the two tests considered in conjunction.

2.5.1. The multiple-choice cloze test

We based our multiple-choice cloze test on two existing open-ended tests that are used as entrance tests at *Interstudie*, a teacher training college in Nijmegen. Mulder (personal communication) suggested two tests which he considered particularly suited for our purposes.

The texts for the two tests had been selected from *La France en direct*, part I (Capelle & Capelle 1969), a coursebook commonly used in French courses for adults, but not in regular secondary education; therefore, the texts were unlikely to be known to any of our subjects. One text deals with the daily routine of going to and returning from work by underground ("Le métro"; henceforth: Métro); the other with French eating and drinking habits ("Comment mangent les français?"; henceforth: Manger).

The texts had been turned into cloze tests by deleting every fifth word, starting with the second sentence. This resulted in 40 blanks for Métro and 42 for Manger. These tests were administered to six SEC-5 students, all of them non-choosers, who met our selection criteria (cf. section 2.5), with the results given in Table 2.2.

As we had hoped, the percentages correct were rather low, allowing us to turn the tests into multiple-choice versions without running too much of a risk of obtaining ceiling effects. As far as they were available, we used the incorrect, inacceptable responses from the administration of the open-ended versions to serve as distractors in the closed-ended ones.

Table 2.2: Percentage correct and reliability (Cronbach's alpha) of the *open-ended cloze tests*.

	Mean	Alpha
Métro	57	.63
Manger	35	.71
Total	46	.79

The (first version of the) multiple-choice (MC) test was administered to a total of 51 subjects, two groups of SEC-5 students (n=15 and n=18), and a group of first-year university students (n=18). Table 2.3 shows the results.

Table 2.3: Reliability of the MC test (version I).

	SEC-5(a) (n=15)	SEC-5(b) (n=18)	UNIV-1 (n=18)	Total (n=51)
Métro	.72	.41	.10	.50
Manger	.46	.60	.40	.54
Total	.73	.68	.42	.67

The percentages correct - not given in the table - were much higher now: 83% for Métro and 77% for Manger, which is quite in line with what one would expect, as is the reduction in reliability (cf. Hinofotis & Snow 1980). Quite unexpected, however, was the enormous variation in reliability across the different groups, in particular for Métro.

We decided to administer the open version of the test to another group in order to obtain more (and different) incorrect responses that could serve as distractors in a second version of the MC tests. This time we used a group of third-year university students (n=18) of different backgrounds in terms of French training. This second administration indeed provided us with ample material to improve the MC version: in all, 34 of the 82 items were changed, 14 in Métro and 20 in Manger. In 28 cases we replaced *one* of the distractors, in the other six we replaced *both*. All the items that were

changed had had a low item-rest correlation (R_{ir}) in the first MC version; 15 even had a negative R_{ir}.

Table 2.4: Reliability of the MC test (version II).

Métro	.72
Manger	.86
Total	.90

The second MC version was administered to a group of third-year university students (n=9) as well; the results are given in Table 2.4. Since the test had now reached an acceptable level of reliability, we decided to use the tests in this form, albeit in combination, rather than just one of the two, which had been our original intention. Another advantage of using two tests rather than one is, of course, that the potential influence of the subject matter of the texts involved is reduced. The second - and final - version of the test is given in Appendix 1.

2.5.2. The listening comprehension test

As a test of listening proficiency we used the listening comprehension test produced by the Dutch National Institute for Educational Measurement (CITO) which is used in most Dutch schools (80-90%, CITO claims) as part of the exam, and which was originally developed by Groot (1976). This test was chosen because it represents the way in which listening proficiency is mostly tested in Dutch schools. Since CITO carefully constructs, pilot-tests and adapts the test before the final exam version is 'published' (see below), this meant we could just buy the test and use it in our investigation.

For practical reasons, we chose the 1984 edition of this test: most of our subjects would not have heard or seen the test in school, since they had either left school or had stopped attending French lessons *before* 1984. There was one exception in this respect: the SEC-6 students who had chosen French; they were tested in the 1987 test session, and attended French lessons right up to that point. In order to avoid a serious bias, we asked the French teachers in both schools to refrain from using the 1984 edition when preparing their students for the listening proficiency exam.

As a rule, the test consists of three parts (A, B and C). Each part consists of a number of text fragments that form a "self-contained argument", so the test manual tells us. Each fragment is followed by a short pause in which the testees have to fill in an MC question about the fragment. When there are three alternatives to choose from, the pause is 22 or 25 seconds long (depending on the text type); when there are only two, it is 8 seconds long. Both questions and answering alternatives are provided in print, and they are all in French. According to the test manual, the aim of the test is to measure the comprehension of spontaneous speech produced by native speakers, speech that is free from too much regional, social or individual colouring, and in which subjects are treated that are relatively non-specialist. Usually, the three parts of the test are three different text types. In the 1984 edition they were: (A) a one-to-one interview about the French apprenticeship system (with 16 questions; 3 alternatives; pauses: 22 seconds); (B) a panel discussion about children's heroes - among them, /ə/-/te:/, normally known as *ET*, /i:/-/ti:/ - (24 questions; 2 alternatives; pauses: 8 seconds), and (C) a short lecture on Christiaan Huygens, a famous 17th-century Dutch scholar (18 questions; 3 alternatives; pauses: 25 seconds). So, there are 58 questions in all; the total running time of the tape is ca. 70 minutes. In line with CITO's recommendations, the testees were allowed a 10-minute break between parts B and C. For readers not quite familiar with this type of test, a couple of examples are given in Appendix 2.

From the 1984 population of the national exam, CITO drew a sample of 926 subjects. Their average score was 42 (out of 58, i.e. 72% correct), and the test reliability (KR20) was .79.

2.5.3. The reading comprehension test

As with the listening comprehension test, we used for reading comprehension the test developed by CITO as part of the official national exam. The only difference with the listening comprehension test is that it is an *obligatory* part of the exam in *all* Dutch secondary schools, and that the test, although carefully constructed, is not pilot-tested.

The test format was originally developed by Gras (1967), who showed that it measured reading proficiency in a very valid and relatively reliable way, compared to the translation, which had been used for that purpose until then. His investigation led to their introduction as *the* final exam for

foreign languages. The test usually consists of five authentic reading passages, mostly taken from prestigious French newspapers, with a total of 50 MC questions more or less evenly distributed among the texts (see Appendix 3 for a sample). As with the listening comprehension test, the questions and the four answering alternatives are in French.

The 1984 edition of the test - which was chosen for the same reasons as the listening comprehension test; cf. section 2.5.2 above - consisted of 5 texts. Text I dealt with the revival of the Classic languages outside the classroom (10 questions); text II dealt with the do-it-yourself boom, and its social consequences (11 questions); text III was an editorial comment on the practice of literary criticism (10 questions); text IV discussed the possible behavioural and social consequences of the introduction of video games (10 questions); and text V dealt with the socio-psychological consequences of studying abroad (9 questions).

From the total population of 15,233 examinees, CITO drew a sample of 1,015 students. A reliability analysis resulted in a KR20 coefficient of .82, and the average score was 39 (77% correct). As a result, the exam was characterized as "reasonable" and "on the easy side", both in the press and in CITO's own evaluation report (Luijten 1984:59).

2.5.4. The phonological subtests

As indicated in section 2.4, we wanted to test phonology for each of the two skills separately, i.e. we constructed a phonology-listening test and a phonology-reading test. The latter would seem to be a *contradictio in terminis*; we did in fact design a phoneme-discrimination test and a variant of the rhyme test, the latter of which tested grapheme-phoneme correspondences.

In both tests minimal pairs were presented, i.e. pairs of words that differ by only *one* phoneme. The phonology-listening (PHO-LI) test was an ABX-like test (cf. Butcher 1976), i.e. triads were presented consisting of two identical words and a minimally contrasting one. In a 'true' ABX-test, subjects are asked whether the third word of the triad (X) sounds like the first (A) or like the second (B). In our test, the identical words were allowed to be in any of the three positions in the triad, i.e. we randomly assigned to each of the items one of the possible orders AAB, ABA, BAA, BBA, BAB, ABB. Each triad was presented twice, with a one-second break between the two presentations. The subjects were instructed to indicate on the

answering sheet, during the two-second break following the item, which of the words were identical. A similar set-up was used by Cross (1982), with one practical difference: he asked his subjects to indicate whether the three words had been pronounced in the same way (S) or differently (D). Instead, we only asked for *same* indications by means of ticks, a set-up that had proved satisfactory in a pilot-study (cf. Van Agt & Wessels 1984). In total we tested 34 phonemic oppositions (see below), and six filler items were added, in which all *three* words were identical (see Appendix 4). The items were presented in a randomized order, and the order of the three words per item was randomized as well. The words were pronounced by a native speaker of standard French, and recorded in the recording studio of the Institute of Phonetics of the University of Nijmegen. We used Studer professional recording equipment and BASF studio-quality audio tape. In the process of the item selection, we based ourselves on the contrastive description of French and Dutch phonology in Kleijn (1977).

The phonology-reading (PHO-RE) test was of the AX-type: subjects were asked to give *same-different* judgements about pairs of words. Since the words were presented in print, we decided that the pairs should differ only minimally in sound, as well as in spelling, e.g. we would include a pair like *peau-beau*, rather than one such as *pot-beau*. In this way we ensured that the response was clearly related to the contrast we wanted to test. As in the PHO-LI test, we included 34 'real' items in the PHO-RE, but also 34 fillers (cf. Appendix 5). We chose equal numbers of target and filler items to ensure a (potentially) equal distribution of *same* and *different* responses. In order to avoid a situation in which the spelling alone would give away the item - viz. one in which the *same* items were spelled identically, and the *different* items were not - the filler items consisted of pairs of homophones. These were based, whenever possible, on the pairs in the target items, e.g. *peau-beau* (target) and *peau-pot* (filler); *sain-zain* (target) and *saine-seine* (filler). Again, the items were presented in a randomized order, and so were the members of the pairs.

The oppositions tested were evenly distributed across the categories 'similar in Dutch' and 'contrasting with Dutch', i.e. 17 in each category. Of the 17 similarity items, eight were consonantal, voiced/voiceless oppositions; nine were vowel oppositions. In fact, we tested six consonantal and six vocal oppositions in different positions (cf. Tables 2.5. and 2.6.): five of the consonantal oppositions were tested in three positions: initially, medially and

Table 2.5: Phonemic oppositions tested: consonants.

	Initially	Medially	Finally
/p/ - /b/	+	+	-
/t/ - /d/	+	+	-
/k/ - /g/	-	-	-
/f/ - /v/	+	+	-
/s/ - /z/	+	+	-
/ʃ/ - /ʒ/	( )	( )	-

'+' = similar in L1 and FL; '-' = contrasting with L1.

Table 2.6: Phonemic oppositions tested: vowels.

	Monosyllabic	Disyllabic-1	Disyllabic-2
/ɑ/ - /ɔ/	+	+	+
/ɑ/ - /ɛ/	+	+	+
/ɔ/ - /ɛ/	+	+	+
/ɑ̃/ - /ɔ̃/	-	-	-
/ɑ̃/ - /ɛ̃/	-	-	-
/ɔ̃/ - /ɛ̃/	-	-	-

'+' = similar in L1 and FL; '-' = contrasting with L1.

finally, and one only in final position, in order to obtain equal numbers of similarity and contrast items. As is indicated in Table 2.5, all the oppositions are contrasting in final position, as a result of syllable-final devoicing ('Auslautverhärtung') in Dutch. The opposition /k/-/g/ is contrasting in any position, because Dutch does not have a phoneme /g/. It only occurs as an allophonic variant of /k/ in some contexts (in some regional varieties more so than in others), and in a few borrowings, such as *goal* and *goulash*. The same holds for the opposition /ʃ/-/ʒ/, which we only tested in final position, where it would be neutralized anyway: /ʒ/ only occurs in loanwords such as *jury* and *beige*. Of the other oppositions, i.e. /p/-/b/, /t/-/d/, /f/-/v/ and /s/-/z/, the latter two are subject to substantial regional variation in Dutch. Speakers from some northern and western parts of the Netherlands tend to

devoice /v/ and /z/ also in initial position, and sometimes even in medial position (see e.g. Gussenhoven 1981). Nevertheless, we decided to categorize these oppositions as 'similar in L1', since devoicing is hardly ever systematic or complete in any one speaker, and since all our informants would be at least receptively aware of the distinction in view of the fact that devoicing is very uncommon in southern varieties of Dutch, as it is in educated varieties in general.

The vowel distinctions are quite straightforward: the oral variants of /ɑ/, /ɔ/ and /ɛ/ are very similar in French and in Dutch, whereas the French nasal variants are generally absent in Dutch, except in a few loanwords such as *engagement* and *genre*. In order to achieve a balance between the consonantal and the vocalic 'parts' of the test, the vowels were also tested in three positions, viz. in a monosyllabic word, in the first syllable of a disyllabic word (Disyllabic-1), and in the second syllable of a disyllabic word (Disyllabic-2).

The actual test words were selected on the basis of their transcription in Warnant's (1964) pronouncing dictionary. Pairs were found by consulting Juilland's (1965) retrograde dictionary, and Warnant's (1973) rhyming dictionary. The complete list of target and filler items is given in Appendices 6 and 7.

Both tests were pilot-tested with 18 subjects - selected in accordance with the criteria given in section 2.6; SEC-5 students who had all dropped French ca. six months before the test administration. The test reliability (Cronbach's alpha) was rather low: .44 for the PHO-LI test (34 items) and .75 for the PHO-RE test (34 items). It should be noted that of the 34 items in each of the two tests, 17 and 13 respectively were zero variance items; they were answered correctly by *all* subjects, and hence did not contribute to the test reliability. In other words, the analysis was carried out as if we were dealing with 17-item and 21-item tests respectively.

As indicated in section 2.5 above, there is an alternative way of defining test reliability in our case, namely by using 'composite scores' - sometimes also referred to as 'super items' - instead of actual item scores. The problem then is, of course, that these tests would consist of only four items each (consonantal-similar, consonantal-contrasting, vocalic-similar and vocalic-contrasting). Therefore, we decided to run a reliability analysis across the *two* tests, i.e. we computed the composite scores for the four subsets of items in either test, and treated these eight scores as if they were items. Defined in this way, the reliability of the phonological test was

.69; quite an acceptable level in view of the fact that it consisted of only eight 'items', and in view of the general considerations mentioned in section 2.5 above.

2.5.5. The lexical subtests

Both lexical subtests had exactly the same format: a French sentence was presented, plus its translation with the target word left out, a test format comparable to the one used by, for example, Macht & Steiner (1983). The only difference between the oral and the written version was the way in which the French sentence was presented: in the lexical listening test (LEX-LI) it was presented on audio tape, in the lexical reading test (LEX-RE) it was presented in print. In the LEX-LI test the pause following each sentence was 10 seconds. In both tests subjects were asked to fill in the missing word in the Dutch translation.

Each of the two tests consisted of 40 items, evenly distributed among the four categories resulting from the combinations between the two dichotomies high vs. low frequency, and cognate vs. non-cognate. All target words were selected from the *Vive le français* glossary, in order to ensure that the words had indeed been learned, or rather, taught at some point in the curriculum. Frequency was based on Savard & Richards (1970), a complete frequency list of the 3200 words included in *Le français fondamental I* and *Le français fondamental II*, the basis for the *Vive le français* glossary. (The list in Savard & Richards 1970 is actually based on a 'utility index' rather than just frequency, but in view of its rather dubious basis - cf. Schils & Reelick 1985 - we preferred frequency as the criterion for our selection.) The high-frequency targets were selected from the first 750 words, the low-frequency ones from between words from 1750 to 3200.

There was one category, however, in which we deviated from this procedure, viz. the low-frequency cognates (for our definition of 'cognates', see below). The procedure described above resulted in too few of these words, and we, therefore, added a number of words that did not occur in the coursebook, but which could be expected to occur in the curriculum anyhow. They included words such as *finances* (Du. "financiën") and *masse* (Du. "massa").

Cognates are usually defined as words which are the same or very similar in spelling and meaning in L1 and L2 (see e.g. Van Els et al. 1984:216-218). One of the studies on cognates cited most is the one by

Hammer & Monod (1976). They defined cognates as words which are comparable in meaning in the two languages - in their case, English and French - and which differ in spelling by one letter only, or which have the same stem but a language specific morphological ending, e.g. Fr. *varier* vs. Eng. "to vary", and Fr. *curieux* vs. Eng. "curious". Our objection against this definition is that words which are very similar in appearance do not qualify as cognates, especially in cases where we are dealing with longer words which still look very much alike, even if they differ by more than one letter, e.g. Fr. *circonstance* and Eng. "circumstance".

Therefore, we widened our definition in two ways. Firstly, we changed 'difference' into 'transformation' (see below); and, secondly, we allowed a maximum of two such 'transformation' differences. A 'transformation' usually meant the same as a 'difference', but it also included wider differences, such as the deletion of more than one letter or the conversion of two letters into one (other) letter. The general idea is best demonstrated by means of a few examples. In the case of *classe*, the Dutch equivalent "klas" may be derived by (1) changing *c* into "k", and (2) deleting the two final letters *-se*. The second transformation is an example of the deletion of more than one letter; in terms of number of letter differences *classe* would *not* qualify as a cognate. An example of the conversion of two letters into one (other) letter is contained in the transformation of *planche* into Dutch plank: (1) *ch* --> "k", and (2) *-e* -->Ø. Here the criterion of two letter differences would also lead to a classification as non-cognate.

In other words, cognates were defined as target words that (1) were very similar in meaning to their Dutch equivalents, and (2) could be deduced from their Dutch equivalents by means of *one* or *two* transformations, in which an end morpheme like *-ion* (Du. "-ie") or *-ique* (Du. "-iek", "-isch") was counted as *one* transformation.

With words intended for the LEX-LI test we counted the number of transformations necessary to deduce the Dutch written form from the French spoken form; for the LEX-RE test we compared the two written forms. To give a few examples: *musique* did not qualify as a cognate for the LEX-LI test, because no transformation at all would be required - the Dutch equivalent "muziek" is, phonemically, almost identical - but it did qualify for the written test: *s* --> "z", *-ique* --> "-iek". Alternatively, *planche* did not qualify for the LEX-LI-test, because three transformations would be required to arrive at the Dutch equivalent "plank"; the written form, however, did meet the criteria, as was explained above.

In order to check, or rather, validate, our frequency criterion, and in order to avoid the inclusion of words that had not been learned/taught at all, we presented the initial selection of 146 words - in alphabetical order - to five teachers of French at one of the two schools we drew our subjects from. They were asked to indicate for each individual word how often if would occur in the first four years of French training. The five scale points were marked (1) never, (2) occasionally, (3) sometimes, (4) regularly, and (5) often.

These judgements were first subjected to a reliability analysis. It turned out that they were very reliable: the ratings of the individual judges all correlated very significantly ($.50 \leq r \leq .73$, $df=144$, $p<.001$), and Cronbach's alpha was .89 ($n=146$, $k=5$). Therefore, we decided to use the sum of the five ratings as a criterion for our selection. Since the grand mean of the ratings was 3.1, we used the total of 15 as the cutting-off point between high- and low-frequency words: words with a total rating over 15 were included in the first category; those with a total rating lower than or equal to 15 in the last category. In addition, the total rating of a low-frequency word had to be at least 8: in this way we ensured that at least three out of five teachers thought the word in question occurred "occasionally" in the curriculum.

As a result of the fact that we did not leave a gap between high and low frequency in evaluating the teachers' ratings ($R_{tot}>15$ and $R_{tot}\leq 15$ respectively), it might have been the case that the two categories were not far enough apart to obtain a frequency effect at all. To find out whether this was the case, we compared the average ratings of two categories (see Table 2.7).

Table 2.7: Average teacher ratings of the high- and low-frequency words ($n=146$).

	High	Low	t	df	Sign
Cognates	3.8	2.6	5.25	67	$p<.001$
Non-cognates	3.9	2.3	8.40	75	$p<.001$

This analysis showed that, for both cognates and non-cognates, the difference between high and low frequency was sufficiently large, viz. a difference of 1.2 and 1.6 scale points respectively. The high-frequency words

were rated on average as "occurring regularly" (4); those of low frequency as "occurring occasionally" (2) to "occurring sometimes" (3). It should be noted that this analysis was carried out on the complete list of 146 words, which means that it still included words that violated our criteria in one way or another. The differences between the words that were actually selected were of course even larger, viz. 1.8 and 2.2 scale points respectively (see Table 2.8).

Table 2.8: Average teacher ratings of the high- and low-frequency words selected for the test (n=80).

	High	Low	t	df	Sign
Cognates	4.2	2.4	10.14	38	p<.001
Non-cognates	4.3	2.1	11.37	38	p<.001

The words that were finally selected as target words are listed in Appendices 6 and 7. The sentences for the listening test were pronounced by the same native speaker, and recorded with the same equipment that was described above (section 2.5.4).

Both tests were pilot-tested with the same 18 subjects described in section 2.5.4 above. The reliability coefficients (Cronbach's alpha) were .68 for the LEX-LI test (14 zero variance items, all 100% correct) and .27 for the LEX-RE test (20 zero variance items). In the same way as with the phonological subtests, we also computed Cronbach's alpha over the eight 'subscores' (four in either test: HI-COG, LO-COG, HI-NOCOG, and LO-NOCOG); defined in this way, the reliability of the subtests was .77.

2.5.6. The morpho-syntactic subtests

The phenomena to be tested in the morpho-syntactic subtests were all selected from Eggermont & Hoekstra's (1975) *Grammaire fondamentale*, the grammar book that goes with Heurlin's (1972) course. They were selected in such a way that they more or less represent the teaching content of the grammar book. Firstly, we only selected the more general rules rather than marginal phenomena or exceptions. Secondly, we determined the number of items in proportion to the amount of space devoted to each

particular aspect in the grammar book, e.g. only one item dealing with adverbs, but 12 dealing with different verb forms.

The test format we chose was multiple-choice blank-filling. The following item is an example:

Voulez-vous . . . votre valise à la gare?

0 *laissez*
0 *laissé*
0 *laisser*

We selected a total of 40 phenomena: 20 phenomena that French grammar shares with Dutch grammar - for example the use of the infinitive in the above example - and 20 phenomena that represent contrasts between the grammars of the two languages, thus representing the factor (absence or presence of) contrast between L1 and FL (see Appendix 8 for the complete list of phenomena tested).

The 'contrast' items tested in fact three types of differences between the two languages:

(1) items which tested phenomena that do not exist at all in Dutch, e.g. the partitive article (cf. Fr. *de la confiture* vs. Du. "jam"), and m/f gender marking in adjectives;
(2) items in which a functional difference is tested that is indicated by means of a formal difference in French, but not so in Dutch, e.g. the difference between predicative adjectives and adverbs (cf. Fr. *lente* - *lentement* vs. Du. "langzaam" in either case), and the use of the subjunctive after - for example - *avoir peur que*;
(3) items which test phenomena that exist in both languages, but with a different distribution. Reflexive verbs, for example, occur in both languages, but some of the verbs which are reflexive in French are not reflexive in Dutch (cf. Fr. *se promener* vs. Du. "wandelen"); both languages have gender and case marking of singular personal pronouns, but Dutch has one oblique form for each gender ("hem", m.; "haar", f.), while French distinguishes between dative and accusative (*lui-le*; *lui-la*), but not between masculine and feminine in datives (*lui-lui*).

The difference between 'similarity' and 'contrast' items was further enhanced by the selection of the two distractors: for the contrast items we chose one or two distractors on the basis of what *Dutch* grammar would

predict. In this process we also used Knibbeler's (1977) report on grammatical errors in Dutch learners' French.

The morpho-syntactic listening test (MS-LI) was constructed with the help of the native speaker and the equipment described before. In order to ensure the production of normal intonation patterns, the *complete* sentences were recorded - rather than sentences-with-a-blank - and the targets were edited out later. They were replaced by a sequence of pause-beep-pause which was 3x300 milliseconds long, irrespective of the length of the fragment that had been removed. It would have been undesirable to make the length of the replacing sequence anyway near that of the removed fragment, firstly, because that would have virtually given away the item in some cases, and, secondly, because some of the removed fragments were so short (e.g. *a*), that a beep of that duration would have been hardly audible. Therefore, a *constant* 900 milliseconds replacing fragment was put in in all cases, also in those cases where nothing had been deleted, i.e. where the correct answer was "Ø". In virtually all cases this operation could be carried out by hand, i.e. by means of the editing facility in the recording studio. In two cases this approach proved impracticable, because we were unable to determine the cut-off point between the words to be edited out and those to be kept in. In those two cases, we kindly used the MOSES speech editing system of the Institute of Phonetics of the University of Nijmegen. Each sentence was presented twice, with an interval of one second, and followed by a 10-second pause. The answering sheet only contained the item numbers plus the three answering alternatives.

In the written test all the material was in print, of course. It tested the same phenomena as the MS-LI test, but in a different context, by means of different words, comparable to the set-up in the phonological subtests, where we tested the same phonemic oppositions in different words. The 40 phenomena were randomized for each test separately.

The pilot test was carried out in the same way as for the other subtests. Measured in the 'traditional' manner, the reliability of the two tests was .72 and .50 for the MS-LI and the MS-RE respectively. Measured in the way described above, by means of the composite scores - in this case only four: similarity and contrast items in either test - Cronbach's alpha reached .80, quite an acceptable level for a four-item test, of course.

2.5.7. The questionnaire: Self-report data

The questionnaire served a number of purposes: (1) to gather information about the subjects' (learning) background which could be relevant for the selection procedure (exam subjects, number of years of Latin instruction, number of 'resits', school career, etc.; cf. also section 2.6 below); (2) to obtain some idea of the subjects' attitudes towards French; (3) to obtain self-report information on their (past and present) proficiency in French.

Three questions were included that were meant to give a rough indication of the subjects' attitude towards French:

8. What do/did you think of the French lessons?

terrible	not nice	reasonable	nice	very nice
0	0	0	0	0

9. What do you think of French as a language?

terrible	not nice	reasonable	nice	very nice
0	0	0	0	0

10a. Do you think you will ever need French again? 0 yes 0 no

10b. If so, for: 0 professional purposes
0 study
0 holiday
0 reading books, papers, etc.
0 something else, viz. ...
(More than one choice allowed).

The self-report questions on proficiency level in French were, on the one hand, three very general questions asking subjects to rate their pronunciation, vocabulary, and grammatical knowledge on a five-point scale marked (1) very bad, (2) bad, (3) reasonable, (4) good, (5) very good (cf. Evers 1980). They were asked to give a rating for the moment of the test administration ("now"), and for the moment they gave up French at school ("then"). In this way we hoped to get an idea of how serious the subjects themselves thought their attrition had been, and maybe also a - very global - indication of which linguistic levels they thought had suffered most. For a discussion of so-called retrospective measurements, see Howard & Dailey (1979), Clark (1982), and Sprangers & Hoogstraten (1988).

Apart from these general questions, there were two so-called *can-do scales*, one for listening proficiency and one for reading proficiency. The scales were developed by Clark (1981), and subsequently used - in slightly adapted versions - in a number of projects, e.g. Gardner et al. (1985, 1987), Clark & Jorden (1984), and De Bot & Lintsen (1986). We used the original scales from Clark (1981), with two changes: (1) we translated the items into Dutch, and (2) as in Gardner et al. (1985), we used a five-point scale with all five points marked.

The point of can-do scales is that a number of realistic situations or actitvities are presented to the subjects, and that they are asked to indicate for each of these situations how well they would be able to perform in them, using a five-point scale: (1) not at all, (2) with extreme difficulty, (3) with a lot of difficulty, (4) with some difficulty, (5) with little or no difficulty. As with the global self-assessments for the three linguistic levels, the subjects were asked to give an indication of their proficiency "then" and "now" (see above).

The listening scale consisted of 11 items ranging from: "Understand simple statements like "Hello", "What's your name?", "Where do you live?", to "On the telephone, understand a Frenchman who is speaking as rapidly and as colloquially as (s)he would to another Frenchman". The reading scale consisted of eight items, also ranging in difficulty level from very basic to highly advanced. The complete scales are given - in English - in Appendix 9.

In the pilot tests we gave the entire questionnaire to the 18 subjects that also took the subtests (see above), plus 13 others: first-year university students who had given up French two and a half years before, i.e. they had the same training in French as the other subjects, but two more years of non-use. In order to get some feedback on the questionnaire, the pilot version had an extra question:

(14) Did you find the questions clear and answerable? If not, please indicate which questions you found difficult, and why (e.g. wording unclear; does not apply to my situation; a particular option was lacking).

Positive comments were made by five subjects, and none at all by 16 others. This means only ten subjects made negative comments: two subjects

found it difficult to remember their school results; the other eight made (one or more) remarks about the self-report questions:

- questions resemble each other too much (n=3);
- difficulties dependent on topic addressed (n=2);
- instructions for "then" judgments unclear (n=1);
- self-reports difficult (n=4).

In all, we thought this was a satisfactory evaluation; the only change we made was in the wording of the instructions for the "then" judgements.

A second aspect we looked at was the reliability of the self-report data. To this end we computed Cronbach's alpha for the two scales, for the momentary ("now") and retrospective ("then") judgements separately. The results of this analysis are given in Table 2.9, which shows that the self-reports were very reliable.

Table 2.9: Reliability of the can-do scales in the pilot test.

	"Now"	"Then"
Listening comprehension	.88	.90
Reading comprehension	.86	.88

An analysis of the item means showed that there was no indication of bottom or ceiling effects: they ranged from 1.5 to 4.7 for listening (grand mean 2.8), and from 1.6 to 4.6 for reading (grand mean 3.1). In other words, both the scales seemed to adequately cover the range of our subjects' proficiency in different situations.

2.6. THE SUBJECTS

In Table 2.10 we present - once more - the six groups, consisting of 25 subjects each, that took part in our investigation. The letters A to F used to designate the groups should be read as follows:

Table 2.10: Design of the research project (The arrows indicate longitudinal measurements).

		Years of non-use				
		0		2		4
Years of training	6	A		B	-->	C
	4	D	-->	E		F

A = students of grade 6 of secondary school (SEC-6) who have had six years of training in French, right up to the moment of testing;
B = second-year university students (UNIV-2) who have had six years of training in French, plus two years of non-use;
C = fourth-year university students (UNIV-4) who have had six years of training in French and four years of non-use;
D = students of grade four of secondary school (SEC-4) who have had four years of training in French, right up to the moment of testing;
E = SEC-6 students with four years of training in French and two years of non-use;
F = UNIV-6 students with four years of training in French and four years of non-use.

Since the developments from group B to C, and from D to E were studied longitudinally, no matching was necessary between them; the other groups, however, had to be selected in such a way that they were as comparable as possible with respect to potentially interfering variables. This, then, applied to groups A, B, D, and F.

Since there is, as yet, no exhaustive descriptive model of FL attrition, we had to make a selection on more or less intuitive grounds, and on the basis of more or less hypothetical accounts, such as Berko-Gleason (1982) and Oxford (1982a, 1982b). In the selection procedure we intended to control the following variables:

- learning career;
- general language ability;
- number of FLs studied after SEC-4;
- training in Latin;

- post-course exposure to French.

Latin instruction was controlled particularly because of the expected transfer on the lexical level. We were aware of the fact that there is a lot of other FL input that might have to be controlled, but as will become clear from subsequent sections, we had considerable trouble finding subjects who met the criteria listed above. The selection of the subjects for the different groups is discussed in the following two sections.

2.6.1. The secondary school students

The secondary school students were all recruited from two schools in Nijmegen, the Elshof College and the Stedelijke Scholengemeenschap. Both schools use the same coursebook, *Vive le français* (Heurlin 1972), and the accompanying grammar book, *Le grammaire fondamental* (Eggermont & Hoekstra 1975). In this way we achieved an important degree of control over the variable *learning career*: all secondary school students had learned French from the same coursebook, at roughly the same rate, with the same grammar book, and the same word lists. Additionally, we tried to get as many students from the 'language stream' (*Atheneum-A*), where two modern FLs are obligatory as exam subjects and Latin is not taught. In this way, we felt, we would control *learning career* even further, and at the same time control *general language ability* to some degree, the *number of foreign languages studied*, and the *amount of Latin training*. In practice, however, we could not apply the two-foreign-languages criterion too strictly without losing too many subjects: we also had to allow some *Atheneum-B* students who chose only one FL in their exam package.

Similarly, we had to allow some training in Latin. Although all our subjects took the *Atheneum* exam, some of them had had some Latin in the lower grades of secondary school. In practice, we allowed a maximum of three years of instruction in Latin. This means that none of our subjects attended any Latin classes after SEC-4; in other words, there could not have been any retro-active transfer from Latin to French for the non-choosers.

Another relevant aspect of *learning career* is the number of resits, i.e. the number of times a student failed to meet the criteria for annual promotion and had to attend the same grade once more. We had intended to use only subjects without any resits, but this also turned out to be impossible in

view of the frequent occurrence of this phenomenon. So we adopted a policy comparable to the one we employed for Latin instruction: we allowed a maximum of two resits, but none after SEC-4. Even then, no less than 19% of the subjects in our final sample had missed annual promotion once or twice.

A further check on *general language ability* was built into the questionnaire: we asked our subjects to indicate their average school marks in SEC-1 through SEC-4 for all modern languages they had taken, i.e. Dutch, English, German, and French. The sum of these was taken as an indication of *general language ability*.

Post-course exposure to French was also probed in the questionnaire: the subjects were asked to indicate whether they (had) used French outside the school. They were also requested to indicate how often this contact had occurred - if at all - and how much actual use of French had occurred. This latter question was included, because we suspected that contacts with France did not necessarily mean that French would be used a lot, a suspicion that was confirmed by the numerous responses saying something like: "I've been to France several times, but I hardly ever spoke French".

As indicated above, we had good reasons to assume that contacts with French would not be frequent. In fact, the information obtained through the questionnaire led to the exclusion of only very few subjects: a few were removed because they had French-speaking relatives whom they met and corresponded with regularly; a few others were removed because they indicated they were film fanatics and frequently watched French-spoken films on Walloon television.

2.6.2. The university students

The university students were selected on the basis of the same criteria as the secondary school students, i.e. our advertisement in the university weekly called for people with an *Atheneum* diploma, preferably with two languages in their exam package, but no Latin. Additionally, we did not allow any language or linguistics students in our sample, because we suspected their pre-occupation with a FL, or with the phenomenon of language in general, might bias their response behaviour.

As with the secondary school students, we had to allow some people with only one FL after SEC-4, some people with one, two, or three years of

Latin instruction, and some people who had attended one or two grades twice (cf. 2.6.1 above).

Finally, we did not succeed in gathering enough university students who had used the *Vive le français* coursebook: as could have been expected on the basis of the national situation (see Kuhlemeier & Van Werkhoven 1984), only about half (52%) of the students had used this coursebook in secondary school. However, another 19% had used the coursebook *On parle français* (Hellström & Johansson 1969), which is in many respects highly comparable to *Vive le français* (see Bijlsma 1976). Another 11% did not remember which coursebook they had been using.

2.7. PROCEDURE

The subjects were tested in two test sessions on two consecutive days. Testing took place in one of the language labs of the Department of Applied Linguistics of the University of Nijmegen, enabling us to use the audio equipment for the presentation of the aural tests, and to ensure that the administration conditions were equivalent for all subjects. The first test session consisted of the following parts:

- a warming-up test (see below);
- the multiple-choice cloze test;
- the phonological subtests;
- the lexical subtests;
- the morpho-syntactic subtests.

Including the breaks between tests, this session took about three-and-a-half hours. The second test session, which took about four hours, consisted of the listening comprehension and the reading comprehension tests. The subjects were paid the usual fee for their participation.

The subjects from SEC-4 only sat through the first session; in other words, they did not take the listening and reading proficiency tests. This decision had been made because these tests are specifically aimed at the cognitive level of SEC-6 students, and we thought it was, therefore, unfair to administer such tests to SEC-4 students.

The warming-up test was included, because we suspected that the subjects who had not used French for two or four years might have to get used to the language again. In other words, we wanted to avoid them from

having a 'cold start', and therefore included the warming-up test before the first actual test. Obviously, we gave the warming-up test to *all* groups. On the other hand, in line with common practice regarding warming-up tests, the test should not be too long, and it should test skills or knowledge comparable to, but sufficiently different from, the target material in the actual tests.

We chose test no. 2022 from the package *Leerdoelgerichte toetsen Frans luistervaardigheid* (CITO 1983). It tests the use of place indications, such as *en face de*, *loin*, and *tout droit*, which did not occur as targets in any of our actual tests. The test consists of eight short question-and-answer exchanges between two people. After each fragment a question is given - in Dutch - with two options. The selection of a listening test had the additional advantage that the subjects could familiarize with the language lab equipment which would also be used for the actual listening tests.

On the basis of the experiences in the pilot tests, all tests were paced in such a way that everyone was able to finish them without any time pressure. The two CITO tests were, of course, used in their official form; in practice, this also meant that everyone had enough time available to complete them at a normal pace.

3. RESULTS

In this chapter we will present all the results of the investigation. First of all, we will check our subject sample for any cohort effects (3.1). In the subsequent paragraphs (3.2 to 3.7) the results of each of the tests will be presented, in the same order as in chapter 2. The test scores were analyzed by means of the 8V programme from the BMDP package, *analysis of variance for equal cell sizes and mixed models*, and the 2V programme from the same package, *analysis of variance and covariance with repeated measures* (see Dixon 1983). The 'mixed model' type of analysis was necessary because the factor Subjects was a so-called 'random factor', whereas all the other factors involved were so-called 'fixed factors' (see, for example, Ferguson 1981:257ff).

Since our design involved a combination of cross-sectional and longitudinal comparisons (cf. section 2.3), there was no straightforward analysis available, however. Any two (interesting) comparisons would inevitably involve one *between*-groups and one *within*-group comparison, because the longitudinal measurements had been made across the second interval of two years for training level 1, and across the first for level 2 (see Table 2.10).

The solution to this problem was to initially carry out an analysis of variance in which only the 'extremes' of our design, i.e. groups A, C, D, and F, were included. This choice was made, because it enabled us to run a straightforward ANOVA in which all comparisons were 'cross-sectional', that is *between*-groups comparisons. In this analysis, the factor Subjects was treated as nested under Education (= no. of years of training in French: 4, 6) and Non-use (= no. of years of non-use following that training: 0, 4). The only potential threat to this cut-out from our design was the 'testing' factor (Campbell & Stanley 1966). All groups represented in this analysis were tested for the first time, except group C: they had also been tested two years earlier, then as group B. We would like to argue, however, that a carry-over effect could hardly have occurred, for the following reasons. Firstly, because the time interval between the two measurements was so long. Secondly, because most of the tests were of the multiple-choice type; in other words, subjects had no idea whether their response was adequate or not, so they had no clue as to what to remember (and possibly look up at home!). As for the other tests, which were not of the MC type, one could

safely argue that, in view of the large number of subjects and the enormous number of test items, the effect of individual, incidental remembrance can not have exceeded the statistically *just notable difference*.

Thus, we were in a position to analyze the development of the test scores across the whole period of non-use of four years, but the analysis would not give us any information on possible differential developments during the first as opposed to the second interval of two years. Neither could it tell us whether the development in a given interval was the same for both training levels or not. For that purpose, a computer programme - a so-called *bootstrap procedure* - was developed which tested any two developments (e.g. intervals 1 and 2 for training level 1, or interval 1 for both training levels) against the null hypothesis of there being no interaction, and which took into account the fact that one of the difference scores was a between-groups difference, and the other a within-group difference. The bootstrap procedure was introduced by Efron (1979) as a technique for estimating a statistic's unknown standard error. Chatterjee (1984:253) characterizes the bootstrap distribution of a statistic as follows:

> "The bootstrap distribution is obtained by replacing the unknown distribution by the empirical distribution of the sampled data and then resampling from the empirical distribution to obtain estimates for the random variable of interest".

In our case, however, the specific problem was not a missing sampling theory, but the estimation of the sampling variance and the standard error associated with the difference between a dependent and an independent difference, as the sampling variances of these two differences are principally heterogeneous. This problem, then, may also be solved by means of a bootstrap approximation (for further details, see Schils 1988).

Whenever we speak of '(non-)significant' effects in connection with the test results, we will be referring to the following cut-off points:

$p \leq .001$:highly significant;
$.001 < p \leq .05$:significant;
$.05 < p \leq .10$:marginally significant;
$p > .10$:not significant.

After all test results, the results of the self-report measures will be presented (3.8). There we will be using the same statistical techniques as with the test scores (see above). Next, we will devote two sections to the relation between different test scores (3.9), and the relation between test scores and self-report data (3.10) and, finally, we will devote separate sections to a comparison between choosers and non-choosers in SEC-4 (3.11), and to the role of attitudes, motivation and use (3.12).

3.1. THE SUBJECT SAMPLE

In section 2.6 we explained how we would have liked to select our subjects. In reality things were much more complicated. Therefore, we will give some indication of the sample that we ended up with here; in particular, we will devote some attention to the control for cohort effects that we built in (cf. section 2.3). This check meant that we tested subjects belonging to groups C and E - the two groups that were measured for the second time at the second time of measurement (cf. Table 2.10) - at the first time of measurement as well. In other words, in these instances we had longitudinal data (from our actual subjects), as well as cross-sectional data (from our control subjects).

Since the control subjects only filled in the questionnaire and the cloze test, we can only compare them with the actual subjects in terms of their general proficiency, and in terms of those variables that were probed in the questionnaire: (self-reported) school marks for modern languages in general and for French in particular, in the lower and higher grades of secondary school; attitudes towards French lessons and French as a language; and potential future utility of French (cf. section 2.5.7). All these variables are listed for both the controls and the actual subjects in Tables 3.1 and 3.2. The former compares two samples of UNIV-4 students: one from the 1985 cohort (the controls, n=28) and one from the 1987 cohort (the actual subjects, n=25); the latter does the same for two comparable samples of SEC-6 students (n=32 and n=25 respectively). The cloze test score should be seen against the background of a maximum score of 82. The questionnaire data are all based on 5-point scales with 1 as the lower end and 5 as the upper end, except for the utility scale, which was a 6-point scale with a minimum of 0). The school marks should be seen in the context of the Dutch marking system, which involves a 10-point scale running from 1 to 10; the critical point is between 5 ("unsatisfactory") and 6 ("satisfactory").

Table 3.1: Comparison of the 1985 and 1987 cohorts: Choosers from UNIV-4.

	1985 (n=28)	1987 (n=25)
Cloze test	71 (4.03)	71 (4.56)
Languages, lower grades	7.1 (0.79)	7.6 (0.57)
French, lower grades	7.0 (0.92)	7.5 (0.76)
Languages, higher grades	7.2 (0.71)	7.3 (0.66)
French, higher grades	6.9 (0.83)	7.1 (0.65)
French lessons	3.3 (0.89)	3.4 (0.70)
French language	4.2 (0.79)	4.0 (0.68)
Future utility	2.3 (1.05)	1.8 (0.87)

The results show that the two cohorts were highly comparable: their general proficiency in French was identical, and of the questionnaire variables only the school marks for French and languages in general in the lower grades showed significant differences ($t(51)=2.29$; $p=.026$ and $t(51)=2.66$; $p=.011$ respectively). In other words, as far as the choosers are concerned, the conclusion seems warranted that there is no cohort effect interfering with our cross-sectional comparison.

The results for the non-choosers from SEC-6 are presented in Table 3.2. First of all, the cloze test score of the controls was higher than that of the actual subjects ($t(55)=2.48$; $p=.016$). Secondly, their school marks in the higher grades were higher, both for languages in general ($t(55)=2.30$; $p=.025$) and for French ($t(55)=2.75$; $p=.008$). And finally, they scored higher on all three attitudinal variables, particularly on the utility variable, where the difference was highly significant ($t(55)=4.58$; $p=.000$). Therefore, we have to conclude that there might be a cohort effect interfering with our results: the data in Table 3.2 suggest that our sample of SEC-6 students is relatively poor compared to the 1985 cohort, from which our (1987) UNIV-2 sample originates.

Since we had indications that our data might be confounded with a cohort effect, we decided to rule it out by submitting the test scores to an

Table 3.2: Comparison of the 1985 and 1987 cohorts: Non-choosers from SEC-6.

	1985 (n=32)	1987 (n=25)
Cloze test	65 (7.81)	60 (7.91)
Languages, lower grades	7.4 (0.83)	7.4 (0.72)
French, lower grades	7.5 (0.97)	7.3 (0.88)
Languages, higher grades	6.9 (0.84)	6.4 (0.77)
French, higher grades	6.9 (0.94)	6.3 (0.69)
French lessons	2.9 (0.88)	2.6 (0.71)
French language	3.9 (0.89)	3.4 (0.91)
Future utility	1.5 (0.84)	0.6 (0.65)

analysis of covariance (ANCOVA): the scores were adjusted for 'general French ability', before they were entered into the analysis of variance. For this purpose we used the (self-reported) average school marks for French in the higher grades of secondary school. However, it turned out that this affected the results only very marginally. (As an example, we have included a complete ANCOVA, as well as the original ANOVA, for one test in Appendix 10.) Therefore, we stuck to the original simpler analysis described in the introductory section to this chapter.

3.2. THE MULTIPLE-CHOICE CLOZE TEST

The multiple-choice cloze test, our test of general (receptive) proficiency in French, consisted of two texts with 40 and 42 blanks respectively, so that the total number of items was 82. The scores on this test are shown in Table 3.3. It does not take a sophisticated analysis to conclude from this table that the test scores remained stable over time. Obviously, the ANOVA only revealed one (highly) significant effect, namely that of Education ($F(1,96)=55.13$; $p<.001$): six years of training does lead to a significantly higher general proficiency than just four years of training.

However, there is another interesting observation to be made in connection with Table 3.3, namely that the standard deviation increases for the

non-choosers as time progresses: the dispersion among these subjects has just about doubled after four years of non-use.

Table 3.3: Mean scores (s.d.) on the MC cloze test (max. score 82).

		Years of non-use		
		0	2	4
Years of training	6	70 (4.31)	69 (5.02)	71 (4.56)
	4	62 (4.50)	60 (7.91)	61 (9.03)

In some of our pilot studies (see, for example, Verkaik & Van der Wijst 1986; Weltens & Van Els 1986) we had also found indications that general proficiency does not suffer from non-use during the period investigated. Quite a few people attributed these results to the fact that we had used a *multiple-choice* format rather than a 'normal' cloze test. In view of this criticism we felt it was advisable to at least include an open-ended version of the same test in the second test round, allowing us to get data on both formats from the four groups in our design that were tested in the second test round. The open-ended version of the test was administered at the very end of the second test session, i.e. as far apart from the MC version as possible, with all the other tests and the one-day break inbetween. The results are shown in Table 3.4.

Table 3.4: Mean scores (s.d.) on the open-ended cloze test (exact-word method; max. score 82).

		Years of non-use		
		0	2	4
Years of training	6	61 (4.86)	--	64 (5.89)
	4	--	48 (7.82)	49 (12.24)

A striking difference between the results in Tables 3.3 and 3.4 is that the gap between the two education levels has widened considerably. In other

words, in its open-ended form the test does seem to have somewhat more discriminatory power (cf. also Hinofotis & Snow 1980). Nevertheless, no attrition was found: the scores remained stable over time, as far as we can tell from the data in Table 3.4.

Additional evidence for the idea that both versions of the test are strongly related is the fact that the scores on the two versions correlated very highly: r=.90 (n=100; p<.001).

In other words, we have found no evidence here contradicting our finding that these subjects' general proficiency in French remains stable over time.

3.3. THE LISTENING COMPREHENSION TEST

The listening comprehension test consisted of 58 multiple-choice questions about three different texts, and was developed by CITO. For reasons explained in section 2.7, this test was not administered to SEC-4 students. Therefore, we were left with an incomplete design, which led us to carry out separate ANOVAs for the choosers and non-choosers respectively. Before going into them, however, we will first present the results (see Table 3.5).

Table 3.5: Mean scores (s.d.) on the listening comprehension test (max. score 58).

		Years of non-use		
		0	2	4
Years of training	6	42 (6.65)	43 (5.79)	45 (6.68)
	4	--	27 (4.96)	32 (5.56)

For both levels of training, the scores tended to increase, but the increase for the choosers was not significant (F(1,48)=2.40; p=.1280). The increase for the non-choosers, however, was highly significant (F(1,48)=13.93; p<.001). Although the fact that we had to run two separate analyses for the choosers and non-choosers prevents us from testing the interaction between Non-use and Education, we may safely conclude that there is such an interaction here, with the choosers increasing only marginally and the non-choosers doing so considerably.

3.4. THE READING COMPREHENSION TEST

The test we used as our reading comprehension test was the standard French exam test for Dutch secondary schools. It consisted of 50 multiple-choice questions about five different argumentative texts taken from prestigious French newspapers. The results of the reading comprehension test are presented in Table 3.6. As with the listening comprehension test, we were left with an incomplete design, leading us to run separate analyses for choosers and non-choosers. The results were quite obvious: for both levels of training the increase was highly significant (for the choosers, $F(1,48)=23.37$; $p<.001$; for the non-choosers, $F(1,48)=18.48$; $p<.001$).

Table 3.6: Mean scores (s.d.) on the reading comprehension test (max. score 50).

		Years of non-use		
		0	2	4
Years of training	6	35 (4.06)	38 (5.61)	41 (4.75)
	4	--	20 (5.98)	28 (6.12)

An aspect of this particular test that we did not mention yet is the fact that it is relatively time-consuming. Normally, students taking this test in their final exams are allowed two-and-a-half hours to complete it. In our test administration, we had set the time limit to two hours, firstly because the French teachers at the two schools we had drawn our secondary school students from had told us that most exam candidates had taken only one to two hours to finish their exam test, and secondly because the 1984 version was considered relatively easy compared to others (cf. Luijten 1984).

Since it took quite some time to take this test, we were in a position to time it fairly accurately without any extra (technical) trouble other than looking at the clock every time a subject indicated (s)he had finished. Thus, we could also analyze the time factor; the results are given in Table 3.7.

Although there were only small differences between the five groups involved - also between choosers and non-choosers, incidentally - we had a closer look at these data, in the sense that we correlated them with the

Table 3.7: Mean amount of time (s.d.) taken to complete the reading comprehension test (minutes).

		Years of non-use		
		0	2	4
Years of training	6	90 (12.79)	94 (13.72)	86 (16.13)
	4	--	90 (12.33)	96 (15.94)

test scores (cf. Table 3.6 above). This led to quite an interesting finding: for the three groups of choosers, the correlation between test score and amount of time required was negative (-.13, -.25 and -.36 respectively) - indicating a tendency for 'better' students to finish earlier, as one would expect them to - but for the non-choosers the correlation was positive (.28 and .39 respectively), indicating a tendency for those who took more time to obtain a higher score. However, since the differences between the groups in terms of the time required were only small, and the correlations only moderate, it may be safely assumed that this relation between time and performance did not interfere with our interpretation of the test results.

3.5. THE PHONOLOGICAL SUBTESTS

As described in section 2.5.4, the phonological subtests consisted of 34 target items and a number of filler items, six in the PHO-LI test and 34 in the PHO-RE test. The analyses described below only concern the target items. These 34 items fall into four categories: contrast-vocalic, contrast-consonantal, similar-vocalic, and similar-consonantal. Therefore, the scores were analyzed by means of a four-way ANOVA with Education (E) and Non-use (N) as subject factors, and Contrast (C) and Vocalic-Consonantal (V) as item factors.

The results of the PHO-LI test, broken down for Education and Non-use, are given in Table 3.8.

For the PHO-LI test, the four-way ANOVA yielded highly significant effects for each of the four main factors. However, there were also three two-way interactions, one significant and two highly significant. They were the following:

Table 3.8: Mean scores (s.d) on the PHO-LI test (max. score 34).

		Years of non-use		
		0	2	4
Years of training	6	32 (1.46)	33 (1.01)	33 (1.10)
	4	31 (2.48)	32 (1.90)	33 (1.44)

- EV: non-choosers scored lower than choosers (main effect for Education), but only on consonantal oppositions;
- NV: the mean scores improved over time (main effect for Non-use), but this was entirely due to the improvement in the discrimination of consonantal oppositions; however, since the scores for the vocalic oppositions approached 100%, the entire interaction may be interpreted as a result of a ceiling effect in the vocalic category;
- CV: contrast oppositions were more difficult than similarity ones (main effect for C), but the interaction showed that this was only true for consonantal oppositions; in other words, *contrasting consonantal* oppositions, such as *trompe-trombe* and *actif-active* appeared to be the most difficult to recognize. A possible explanation for their particular difficulty may be that these oppositions do exist in Dutch, but are neutralized in these positions. In other words, compared to the contrasting vocalic oppositions investigated, which do not occur at all in Dutch, they might be perceptually less salient.

One should, however, also interpret these effects against the background of the extremely high overall mean scores, i.e. the effects may well be (highly) significant, but even the lowest-scoring categories still have scores of 80% or more.

The results of the PHO-RE test are presented in Table 3.9. Again, the percentages correct were extremely high. Nevertheless, here also there were quite a few effects that reached statistical significance. Interestingly enough, all seven effects of the PHO-LI test discussed above also occurred in the PHO-RE test, and the cell means suggested exactly the same interpretations, but there was an additional effect that made a major difference:

Table 3.9: Mean scores (s.d.) on the PHO-RE test (max. score 34).

		Years of non-use		
		0	2	4
Years of training	6	30 (2.97)	32 (1.95)	32 (1.87)
	4	29 (2.97)	29 (3.55)	31 (2.80)

there was a highly significant three-way interaction between Non-use, Contrast and Vocalic-Consonantal ($F(1,96)=17.11$; $p<.001$). Inspection of the cell means showed that this interaction could be interpreted as indicating that the increase of the scores was mainly due to the improvement in the category contrast-consonantal (see Table 3.10) - i.e. the category mentioned above, of the type *trompe-trombe* and *actif-active*.

Table 3.10: Interaction between Non-use, Contrast and Vocalic-Consonantal in the PHO-RE test (%).

	Consonantal			Vocalic		
Non-use	0	2	4	0	2	4
Similarity	88	92	94	96	96	98
Contrast	68	80	89	96	94	95

However, as we noted before in connection with the interactions in the PHO-LI data discussed above, this interaction might again be largely due to the fact that the other categories of items already scored 88% or more to begin with, thus leaving hardly any room for improvement.

Although the scores on both phonological subtests improved over time, we nevertheless had a look at the scores on the level of the individual items. It turned out that both tests contained a number of items that exhibited attrition - in addition to a much larger number that exhibited improvement, of course. In the aural version there were four such items, but they were different ones for the choosers and non-choosers.

The results of the written version, on the other hand, were much more systematic in this respect. As might be expected on the basis of the results of the ANOVA discussed above, after four years of non-use both choosers and non-choosers showed most improvement on contrasting consonantal oppositions, such as /k/-/g/ in medial or final position, and /f/-/v/ in final position, as in *écoute-égoutte*, *brique-brigue*, and *neuf-neuve* respectively.

Moreover, there was also some systematicity in terms of attrition: the choosers and non-choosers did less well on six and eight of the 34 items respectively, and the two lists had three items in common. These three items tested the (contrasting vocalic) opposition between the nasal vowels /ɑ̃/ and /ɔ̃/ in three different contexts: *en-on*, *tendu-tondu*, and *répand-répond* (cf. section 2.5.4). Apparently, this particular opposition is lost by some of the informants, at least when the stimuli are presented in writing. When presented aurally, the opposition is not (yet?) as problematic: only the PHO-LI item equivalent to *répand-répond*, namely *néant-néon*, occurred on the list of attrition items; in addition, it only occurred on the list for the non-choosers.

3.6. THE LEXICAL SUBTESTS

Each of the lexical subtests consisted of 40 items evenly distributed across the four item categories: HI-COG, LO-COG, HI-NOCOG, and LO-NOCOG. Therefore, we again carried out a four-way ANOVA with Education (E) and Non-use (N) as subject factors, and Frequency (F) and Cognate (C) as item factors.

Table 3.11: Mean scores (s.d.) on the LEX-LI test (max. score 40).

		Years of non-use		
		0	2	4
Years of training	6	35 (1.94)	35 (1.19)	36 (1.50)
	4	29 (3.49)	30 (3.16)	31 (2.82)

The analysis of the LEX-LI test (Table 3.11) yielded highly significant effects of Education, Frequency and Cognate, and highly significant

two-way interactions between these factors, viz. EF, EC, and CF. However, there was also a highly significant three-way interaction between these three factors. This interaction is presented in Table 3.12.

Table 3.12: Interaction between Education, Frequency and Cognate in the LEX-LI test (max. score 10).

	Cognate		Non-cognate	
Frequency	High	Low	High	Low
Choosers	9.9	9.8	9.8	5.5
Non-choosers	9.6	9.3	8.8	2.6

Although a Scheffé test for *a posteriori* contrasts showed that the difference between choosers and non-choosers was significant in all four categories except "Cognate-High", the means in Table 3.12 show that this difference was mainly due to the large difference in the scores for the non-cognates of low-frequency obtained by the two groups. The overall mean difference between the groups was 4.8 items (35.0 vs. 30.2), and this one subcategory of items alone accounted for a difference of 2.9 items (5.5 vs. 2.6).

The most surprising result of this analysis, however, was a significant effect of non-use ($F(1,96)=6.27$; $p=.0140$) in the direction opposite to what one would expect: it suggested an *increase* in vocabulary knowledge over time. As can be seen from Table 3.11 above, this was particularly true for the non-choosers (EN: $F(1,96)=2.96$; $p=.0884$). On the other hand, this interaction may have been blown up as a result of a possible ceiling effect in the case of the choosers, who scored 88% correct to begin with.

The LEX-RE test rendered the results shown in Table 3.13. As with the LEX-LI test, the ANOVA revealed highly significant effects of Education, Frequency and Cognate, and highly significant interactions between these three (EC, EF, CF), but again they could all be interpreted in the light of the highly significant three-way interaction ECF ($F(1,96)=35.14$; $p<.001$): all of these effects were largely due to the fact that the choosers were particularly superior on low-frequency non-cognates. Of the overall mean difference of 3.4 items between the two training levels, the LO-NOCOG category alone contributed 2.6 items.

Table 3.13: Mean scores (s.d.) on the LEX-RE test (max. score 40).

		Years of non-use		
		0	2	4
Years of training	6	36 (2.07)	35 (2.36)	35 (2.16)
	4	33 (2.53)	32 (1.93)	31 (2.60)

An important difference with the LEX-LI test was the occurrence of a significant effect of Non-use ($F(1,96)=7.18$; $p=.0087$) in the opposite direction, indicating loss of vocabulary. The presence of a significant interaction between Non-use and Cognate ($F(1,96)=6.21$; $p=.0144$) indicated that cognates are less susceptible to loss than non-cognates: on average, cognates decreased only from 19.8 (out of 20) to 19.7, whereas non-cognates decreased from 14.5 to 13.3.

Although the actual loss was only relatively small, we nevertheless considered it useful to carry out an analysis on the level of individual items - i.e. individual words. What we did was to sort the items according to their difference scores - i.e. the amount of attrition or gain - after four years of non-use. The two tests and the two training levels (choosers and non-choosers) were kept separate, with the result that we obtained four lists. For the LEX-LI test, the two lists showed very little resemblance: with the choosers five items exhibited attrition, with the non-choosers seven did; but none of the items occurred on both lists. In fact, the only resemblance between the two was the fact that they both contained a relatively high proportion of LO-NOCOGs: with the choosers three out of five (*virgule*, *tiers*, *profond*), with the non-choosers three out of seven (*reproche*, *outil*, *barrage*).

The two lists for the LEX-RE test, on the other hand, showed remarkable resemblances: they contained eight and 18 attrition items respectively, of which no less than seven occurred in both lists: two HI-NOCOGs (*bateau*, *repas*), and five LO-NOCOGs (*évier*, *mouche*, *vigne*, *caprice*, *grève*). This fits in very well, of course, with the results of the ANOVAs, which showed that (for both training levels) non-cognates exhibited more attrition than cognates. The high proportion of *low-frequency* non-cognates in the lists of 'attrition items' furthermore suggests that these words are particularly

susceptible to attrition. This latter finding is similar to what we found in one of our pilot studies on vocabulary attrition (reported in Verkaik & Van der Wijst 1986).

3.7. THE MORPHO-SYNTACTIC SUBTESTS

The morpho-syntactic subtests, MS-LI and MS-RE, consisted of 40 items each. Of those 40 items, 20 tested grammatical phenomena that were comparable in both French and Dutch: the 'similarity' items; the other 20 tested contrasts between French and Dutch grammar: the 'contrast' items. Table 3.14 shows the overall results for the MS-LI test.

Table 3.14: Mean scores (s.d.) on the MS-LI test (max. score 40).

		Years of non-use		
		0	2	4
Years of training	6	32 (3.84)	28 (4.24)	29 (3.89)
	4	24 (3.93)	20 (4.25)	21 (4.06)

The ANOVA yielded three highly significant main effects: Education ($F(1,96)=96.92$; $p<.001$), Non-use ($F(1,96)=13.61$; $p<.001$), and Item type ($F(1,96)=26.57$; $p<.001$). Inspection of the cell means showed that all three effects were in the expected direction: the choosers outscored the non-choosers; the period of non-use negatively affected grammatical competence; contrast items were more difficult than similarity items (12.1 vs. 13.4 out of 20 on average).

The bootstrap procedure (cf. the introductory section to the present chapter) confirmed what the means already suggested: the attrition was of comparable size for each of the two training levels, and in either case it occurred in the *first* period of two years.

However, the ANOVA also yielded one significant interaction, namely between Non-use and Item type ($F(1,96)=9.66$; $p=.0025$). This interaction is represented in Table 3.15. The results show that, although both categories decrease (=main effect of Non-use), the contrast items decrease more sharply; hence the interaction between Non-use and Item type.

Table 3.15: Interaction between Non-use and Item type in the MS-LI test (max. score 20).

	Years of non-use		
	0	2	4
Similarity	14.1	12.7	13.4
Contrast	13.6	11.2	11.3

Table 3.16: Mean scores (s.d.) on the MS-RE test (max. score 40).

		Years of non-use		
		0	2	4
Years of	6	32 (4.46)	29 (4.54)	29 (3.86)
training	4	26 (3.54)	22 (2.97)	22 (4.57)

The results for the written equivalent of the MS-LI test, the MS-RE test, are shown in Table 3.16. The ANOVA for this test yielded almost the same results as the one for the MS-LI test: three highly significant effects, for Education ($F(1,96)=49.73$; $p<.001$), Non-use ($F(1,96)=18.82$; $p<.001$), and Item type ($F(1,96)=92.71$; $p<.001$), and a (marginally significant) interaction between Non-use and Item type ($F(1,96)=3.65$; $p=.0591$). All differences were in the same direction as with the previous test, i.e. the choosers outscored the non-choosers, non-use negatively affected grammatical competence, and contrast items were more difficult than similarity items (12.1 vs. 14.6 out of 20 on average). The bootstrap procedure also rendered the same results as with the MS-LI test: the attrition was of comparable size for either training level, and in both cases it occurred in the first interval of two years.

The interaction between Non-use and Item type also pointed in the same direction as with the MS-LI test: there tended to be more loss on the contrast items than on the similarity ones (see Table 3.17).

Table 3.17: Interaction between Non-use and Item type in the MS-RE test (max. score 20).

	Years of non-use		
	0	2	4
Similarity	15.6	14.1	14.3
Contrast	13.6	11.3	11.3

In section 2.5.6 we explained that the category of contrasting items did, in fact, consist of three sub-classes: (i) items testing phenomena that are entirely absent in Dutch (n=5); (ii) items testing functional differences that are indicated formally in French, but not so in Dutch (n=8), and (iii) items testing phenomena with a different distribution in the two languages (n=7). We also ran an analysis with these *four* categories of items: similarity items and contrast items of categories (i) through (iii). Since the MS-LI and MS-RE produced such parallel results, this analysis was carried out over the pooled data for both tests. Because of the unequal sizes of the categories, the scores were transformed into percentages correct.

The analysis rendered the effects that could be expected on the basis of the ANOVAs discussed above: highly significant main effects for Education, Non-use, and Item type, but also a significant interaction between Non-use and Item type ($F(3,288)=3.09$; $p=.0275$). The cell means showed that this interaction was not only due to the difference between similarity and contrast items in general (see above), but that there were also marked differences between the three sub-classes of contrast items (see Table 3.18).

The results showed that one sub-class of contrast items, the 'distribution differences', dropped off only marginally, as did the category of similarity items; the heavy losses were in the two remaining sub-classes: the 'absent' and the 'form/function' classes. The latter sub-class is in fact most remarkable, in the sense that it initially scored as high as the similarity items, and at the same time underwent the steepest fall as well.

We would like to add that these results are quite unexpected in view of the interference theory (cf. section 1.4).

The next step was to look at the MS tests *at the item level*. The first

Table 3.18: Interaction between Non-use and Item type (4 levels) in the MS tests (data pooled across MS-LI and MS-RE; % correct).

	Years of non-use		
	0	4	loss
Contrast (1): absent	53.8	43.6	10.2
Contrast (2): form/function	76.9	63.1	13.8
Contrast (3): distribution	57.4	50.7	6.7
Similarity	74.4	69.5	4.9

possibility we considered was that the attrition was simply a function of the difficulty level at the end of training, i.e. the higher the initial score, the higher the loss. That this was certainly not the case already follows from Table 3.18; it was confirmed by the results of the second analysis, discussed below (cf. Table 3.19).

The second analysis we carried out was to look for items that exhibited heavy and systematic attrition. For that purpose, we produced lists of the items sorted on the basis of the amount of attrition. We obtained four lists: choosers and non-choosers on either test. We then identified the items that occurred high on all four lists, i.e. items whose percentage correct decreased by at least 10% across the total interval of 4 years. The result of this procedure was a group of four items (nrs. 2, 6, 10, and 14 from the list in Appendix 8). In order of the 'heaviness' of the attrition, they were the following:

- item (6): relative pronoun, subject/object distinction (*qui-que*);
- item (10): indefinite pronoun, subject/object distinction (*ce qui-ce que*);
- item (14): demonstrative pronoun, singular (*ce-cet-cette*);
- item (2): personal pronoun, direct/indirect object plural (*les-leur*).

Table 3.19 shows the initial difficulty levels - or proportion correct - and the amount of attrition for each of these four items, broken down for the two training levels (Choosers/Non-Choosers) and the two tests. Items (6), (10), and (14) are contrastive items; item (2) is a similarity item, but a special one, because the corresponding distinction in (standard) Dutch, *hen-hun*, is subject to considerable variation at the moment, and one could argue that -

Table 3.19: Items exhibiting heavy and consistent attrition.

	MS-LI			MS-RE			Total
	Diff. Level	Loss C	Loss NC	Diff. Level	Loss C	Loss NC	Mean Loss
(6)	.62	.40	.44	.62	.32	.52	.42
(10)	.46	.44	.28	.58	.44	.24	.35
(14)	.90	.24	.16	.98	.16	.16	.18
(2)	.44	.12	.16	.96	.12	.12	.13

especially for our adolescent subjects - the distinction is non-existent, or at least problematic in Dutch as well (see e.g. Jaspaert et al. 1986:44). In other words, these data form another confirmation of the idea that the greatest problems are to be found in the category of the contrastive items.

In addition, these results confirm the conclusion drawn in one of our pilot studies (Grendel & Poppe 1986; Weltens et al. 1986b) that the French pronominal system, especially in its contrastive aspects, is in particular subject to attrition.

3.8. THE QUESTIONNAIRE: SELF-REPORT DATA

As explained in section 2.5.7, we asked our subjects to judge their present proficiency in French on a number of aspects: can-do scales for listening comprehension (11 items) and reading comprehension (8 items), and a global self-assessment of pronunciation, vocabulary, and grammar, *at the moment of the test administration.* Where applicable, we also asked them to indicate their proficiency on each of these aspects *at the moment when they ended their French training.* For the choosers this meant at the very end of secondary school, i.e. at the end of SEC-6; for the non-choosers it meant after SEC-4. We will refer to these two types of self-reports as "momentary" and "retrospective" judgements respectively; in tables we will also use the designations "now" and "then".

In order to get some idea of the validity of the judgements, we did two things. Firstly, we correlated the (momentary) self-reports with the comparable test results of that same moment. Secondly, we compared the momentary with the retrospective judgements applying to the same point in

time, i.e. we compared the momentary self-judgements of the subjects who were at the end of their French training, with the retrospective judgements of the subjects who had gone through a period of non-use but were also asked to rate their proficiency at the end of their training. Table 3.20 contains the results of the first analysis; note that we computed a total test score for phonology, lexicon and grammar by adding up the scores on the oral and the written versions of the tests, and that the self-assessments of listening and reading comprehension were entered into the analysis as *mean scores* on the respective can-do scales.

Table 3.20: Correlations between (momentary) self-reports and actual test scores.

Listening comprehension	.47	(n=125; p<.001)
Reading comprehension	.52	(n=125; p<.001)
Pronunciation	.20	(n=150; p<.01)
Vocabulary	.32	(n=150; p<.001)
Grammar	.54	(n=150; p<.001)

We found (highly) significant correlations in all cases, although they were not really high. On the other hand, the correlations are quite satisfactory, given what other people have reported on correlations between self-judgements and performance measures (see e.g. Oskarsson 1984; Sprangers & Hoogstraten 1988). In addition, we should take into account that the tests with the lowest correlations, phonology and vocabulary, were relatively easy, with the result that the so-called *restriction-of-range effect* probably occurred: the fact that the scores showed relatively little variation may have had a depressing effect on the correlations involved.

The results of the second analysis are presented in Table 3.21. It gives the mean self-judgements for each of the six groups, the momentary self-judgements in columns 1 and 4; the corresponding retrospective judgements in columns 2 and 3, and 5 and 6 respectively. The retrospective judgements were compared with the equivalent momentary judgements by means of t-tests (two-tailed); the asterisks denote significant (*: p<.01; **: p<.001) deviations.

On the whole, we feel that Table 3.21 presents quite a reassuring picture:

Table 3.21: A comparison of momentary ("now") and retrospective ("then") self-judgements.

	Choosers			Non-choosers		
	SEC-6 (now)	UNIV-2 (then)	UNIV-4 (then)	SEC-4 (now)	SEC-6 (then)	UNIV-2 (then)
Listening	3.3	3.2	3.5	2.7	2.8	3.1**
Reading	3.8	3.7	3.9	3.0	3.1	3.3
Pronunciation	3.4	3.5	3.6	2.9	2.9	3.5*
Vocabulary	3.4	3.7	3.8	2.9	3.2	3.2
Grammar	3.9	3.6	3.8	3.2	3.2	3.3

of the 20 t-tests only two indicated significant differences, both relating to the retrospective judgements of the non-choosers from UNIV-2, i.e. looking back four years in time, to the moment when they dropped French at the end of SEC-4. This result is very satisfying indeed if we compare the results reported by Sprangers & Hoogstraten (1988). They quote a number of investigations that have demonstrated the occurrence of a so-called "response shift bias": "(...) educational training interventions can change subjects' understanding or awareness of the target concept and the estimation of their posttraining level of functioning with respect to that concept" (p. 148). Extrapolated to our design, this would have meant a re-evaluation of one's French proficiency at the end of the training (= "the target concept") due to the experience gained in the period of non-use (= the "educational training intervention"). However, as noted above, there was hardly any sign of such an effect.

Another interesting aspect of our retrospective ratings is the fact that the choosers from UNIV-2 and UNIV-4 are the same people. At both test occasions they were asked to retrospectively judge their French proficiency at the end of their school career; in other words, in 1985 they were asked to look back two years, in 1987 four years. In line with what Howard & Dailey (1979) and Sprangers & Hoogstraten (1988) report, this re-administration does not lead to a shift in the ratings, in our case not even across a period as long as two years.

To sum up, then, the conclusion seems warranted that retrospective self-ratings of performance are both reliable (cf. section 2.5.7, Table 2.9) and - to a certain degree - valid indicators of actual performance. Moreover, they are, of course, indicators of the subjects' *perception* of their proficiency, and as such represent very useful information in their own right.

Our next analysis was an ANOVA of the momentary self-ratings along the same lines as described in the opening section to the present chapter, i.e. we analyzed the end-points of our design (groups A, C, D, F). Table 3.22 contains the mean scores for the can-do scale on listening comprehension.

Table 3.22: Mean scores (s.d.) on the can-do scale for listening comprehension.

		Years of non-use		
		0	2	4
Years of training	6	3.3 (0.67)	2.7 (0.61)	2.6 (0.64)
	4	2.7 (0.38)	2.1 (0.42)	2.4 (0.50)

The ANOVA yielded highly significant main effects of Education ($F(1,96)=16.68$; $p<.001$) and Non-use ($F(1,96)=17.77$; $p<.001$), and a marginally significant interaction between the two ($F(1,96)=3.93$; $p=.0502$). The figures in Table 3.22 show that choosers (rightly) have a higher esteem of their listening comprehension than non-choosers, and that the scores decrease across a period of four years. The interaction, which suggests that choosers find that their proficiency decreases more than non-choosers do, is due to the fact that the non-choosers from UNIV-2 overrated their proficiency somewhat - a finding that is consistent with what we saw in Table 3.21: they also tended to overrate their original proficiency level.

The can-do scale for reading comprehension yielded the results given in Table 3.23. As with listening comprehension, the ANOVA produced highly significant main effects of Education ($F(1,96)=22.39$; $p<.001$) and Non-use ($F(1,96)=24.85$; $p<.001$), and the direction of the effects was also the same: choosers (rightly, again) give higher self-judgements than non-choosers, and both training levels report heavy attrition over time.

Table 3.23: Mean scores (s.d.) on the can-do scale for reading comprehension.

		Years of non-use		
		0	2	4
Years of training	6	3.8 (0.57)	3.1 (0.56)	3.0 (0.62)
	4	3.0 (0.57)	2.3 (0.59)	2.6 (0.60)

In connection with the can-do scales for speaking and understanding, Gardner et al. (1985) noted that "There was not any significant loss on skills that had been overlearned or on skills for which the original level of competence was low. Instead, students demonstrated attrition on medium-level skills" (p. 529). We had already analyzed the data from our pilot study for this possibility, but we had to conclude there that "our subjects ... reported attrition ... on *all* items in *both* scales [i.e listening and reading comprehension]" (Weltens & Van Els 1986:211). A similar picture emerged from the present data: (smaller or larger) losses were reported all across the can-do scales. In addition, there now was a tendency for the highest losses to occur with *higher*-level rather than *medium*-level skills.

Table 3.24: Mean self-reports (s.d.) for pronunciation.

		Years of non-use		
		0	2	4
Years of training	6	3.4 (0.82)	2.8 (0.90)	2.8 (0.85)
	4	2.9 (0.70)	2.5 (1.00)	2.7 (0.84)

In addition to the can-do scales, we also collected general self-reports on pronunciation, vocabulary, and grammar (cf. section 2.5.7). Table 3.24 gives the results for pronunciation. The analysis of variance showed that the effect of Education was marginally significant ($F(1,96)=3.47$; $p=.0655$) and that the effect of Non-use was significant ($F(1,96)=5.57$; $p=.0203$). In other words, there was only a small difference between the training levels,

and the subjects from both levels reported a deterioration over time.

Table 3.25: Mean self-reports (s.d.) for vocabulary.

		Years of non-use		
		0	2	4
Years of training	6	3.4 (0.96)	2.6 (0.58)	2.6 (0.58)
	4	2.9 (0.57)	2.1 (0.70)	2.4 (1.04)

The results for vocabulary are presented in Table 3.25. The ANOVA revealed a significant effect of Education ($F(1,96)=5.44$; $p=.0218$), and a highly significant effect of Non-use ($F(1,96)=18.44$; $p<.001$). Again, the difference between the training levels was relatively small, but the reported decrease was larger here.

Table 3.26: Mean self-reports (s.d.) for grammar.

		Years of non-use		
		0	2	4
Years of training	6	3.9 (0.64)	2.4 (0.71)	2.2 (0.83)
	4	3.2 (0.82)	1.8 (0.62)	1.9 (0.95)

As far as grammar is concerned, the relevant figures are in Table 3.26. The self-reports for grammar showed a significant effect of Education ($F(1,96)=10.10$; $p=.002$), and a highly significant effect of Non-use ($F(1,96)=81.83$; $p<.001$), confirming the suggestion from the means in Table 3.26, which exhibit the largest drop of all self-reports.

3.9. THE RELATION BETWEEN TEST SCORES

Out of the many ways available for comparing the scores on separate tests, we chose two for our present purposes, viz. Pearson correlations and a factor analysis.

As far as correlations are concerned, we would expect that our general proficiency test correlated positively with all other tests, notably with the more global ones, LC and RC, which in turn should intercorrelate highly as well. Moreover, we would expect high positive correlations between the oral and written versions of our subtests, since they had been constructed in such a way that the items were equivalent to a certain degree. These correlations are presented in Table 3.27. (A complete correlation matrix is included as Appendix 11.)

Table 3.27: Correlations between test scores.

Cloze x LC	.70	(n=125; p<.001)
Cloze x RC	.71	(n=125; p<.001)
LC x RC	.83	(n=125; p<.001)
PHO-LI x PHO-RE	.48	(n=150; p<.01)
LEX-LI x LEX-RE	.61	(n=150; p<.001)
MS-LI x MS-RE	.82	(n=150; p<.001)

As we had expected, all correlations shown in Table 3.27 were high, except those between the oral and written versions of the phonological and lexical subtests, which were only moderate. But, as already indicated in section 3.8 above, these were the tests where the *restriction-of-range effect* may have occurred; hence probably the somewhat lower correlations.

The second analysis we carried out was a factor analysis, into which we entered all test scores for the four end-points of our design, i.e. groups A, C, D, and F from Table 2.1. The only tests that could not be used in this analysis were the LC and RC tests, since they had not been administered to group D (see section 2.7 for further details).

The principal components analysis extracted two factors from the data: the first with an eigenvalue of 3.69, accounting for 52.7% of all the variance, and the second with an eigenvalue of 1.39, accounting for another 19.8% of the variance; the total proportion of the variance explained by these two factors was, therefore, 72.5%. After varimax rotation, the factor loadings shown in Table 3.28 emerged. In other words, a clear two-factor structure emerged, with the cloze test, the morpho-syntactic and lexical tests loading

Table 3.28: Factor loadings of all test scores (after varimax rotation).

	FACTOR 1	FACTOR 2
MS-LI	.85	.22
Cloze	.83	.27
MS-RE	.83	-.01
LEX-RE	.82	-.01
LEX-LI	.68	.50
PHO-LI	.12	.86
PHO-RE	.07	.85

on the first factor, and the phonological tests loading on the second. This suggests that the former may be interpreted as a 'general proficiency' factor, while the latter may be interpreted as a 'phonology factor'. This interpretation of the first factor was supported by the fact that both the listening proficiency and the reading proficiency tests correlated quite highly with the scores on factor 1 (r=.69 and .70 resp. for the 75 subjects out of the total of 100 who also took those tests), but not with factor 2 (r=.20 and .21 resp.). On the other hand, one could just as well defend labelling it 'vocabulary plus grammar', since the cloze test in particular has been characterized as a test of, primarily, vocabulary and grammar (see e.g. Jonz 1988). Similarly, both the LC and the RC tests cannot be denied some degree of reliance on the same two linguistic levels either.

The next step was to make a two-dimensional plot of the subjects' scores on factors 1 and 2, with the result shown in figure 3.1.

The figure shows a remarkable separation of the four groups, and a very interesting location of the four group centroids. Keeping in mind that the X-axis represents the 'general' - or 'vocabulary plus grammar - factor, and the Y-axis the 'phonology' factor, we can note the following:

- there is a clear difference in level between A and D on both factors;
- the period of non-use results in a *decrease* for the first factor - visualized as the westward movement from A to C and from D to F - which is of comparable size for both training levels;

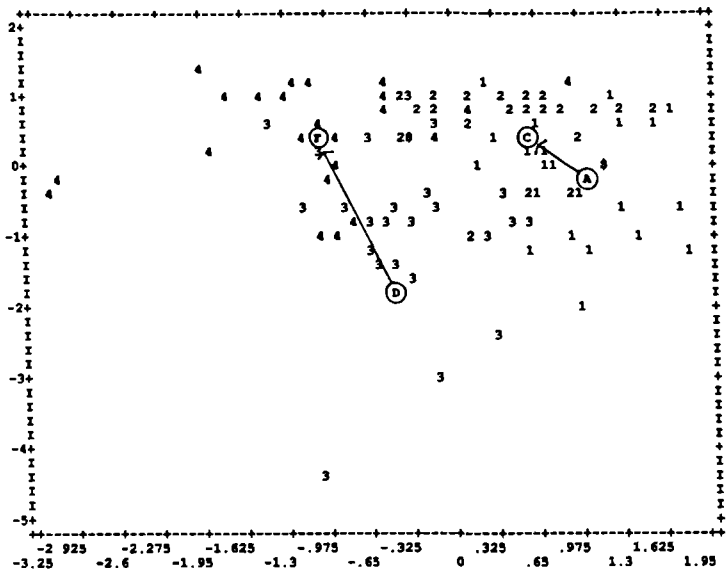

Fig. 3.1: Two-dimensional plot of the subjects' scores on factors 1 and 2. (1 = Subject from group A, i.e. SEC-6 choosers; 2 = Subject from group C, i.e. UNIV-4 choosers; 3 = Subject from group D, i.e. SEC-4 non-choosers; 4 = Subject from group F, i.e. UNIV-2 non-choosers; $ = Multiple occurrences; A-F are the centroids for groups A-F).

- the period of non-use results in an *increase* for the second factor - represented in the northward movement from A to C and from D to F - which is much larger in the latter case, i.e. in the case of the non-choosers, but this is probably the result of a ceiling effect in the phonological test scores (cf. section 3.5 above).

Since the first factor represented a skill that exhibited some attrition, it appeared that its interpretation as 'vocabulary plus grammar' was in fact more likely, in view of the results of the tests contributing to this factor.

3.10. THE RELATION BETWEEN TEST SCORES AND SELF-REPORTS

In section 3.8 we have already seen that the self-report data correlated reasonably well with the actual test scores (cf. Table 3.20). However, when we take into account the development over time, the fit was much less good. The subjects reported heavy losses - of comparable size with each of

the two training levels - for listening and reading comprehension, while the test scores improved over time. Similarly, the phonological tests improved, while the subjects indicated a small decrease. Vocabulary was wrong as well, in the sense that the self-reports indicated heavy attrition, but the LEX-LI test showed a slight improvement and the LEX-RE only a slight loss.

In fact, the only point where test scores and self-reports did nicely match was grammar: both indicated serious attrition, independent of the education level, occurring in the first two years of non-use following training.

Another aspect that came out quite well was the education level. All self-reports exhibited significant effects of education, except those for pronunciation, which were only marginally significant. Exactly the same was found in the test scores: highly significant effects in all tests, except in the two phonological subtests, where they were significant and marginally significant respectively.

3.11. CHOOSERS AND NON-CHOOSERS IN SEC-4

At the time when we tested the SEC-4 students - April/May 1985 - it was not quite the end of the school year, and as a result the students were not yet certain whether they would actually choose French as one of their seven exam subjects; obviously, this choice partly depends on their final marks for French. In order to guarantee that we would be left with enough non-choosers, we tested 80 students, rather than 40 - which we estimated would be enough to leave at least 25 after two years. A second reason for doing this was the fact that some people believed that our two training levels in fact represented a distinction between those who are good at languages and those who are not, respectively. In order to be able to comment upon this presupposition, we had to have data on both groups. Therefore, the choosers from SEC-4 were given the same battery of tests as the non-choosers, i.e. all instruments except the LC and RC tests (cf. section 2.7).

From this original group of 80 subjects, we selected those 50 who were retested in 1987, then as members (choosers and non-choosers) of SEC-6. We then compared the two groups, with the results shown in Tables 3.29 and 3.30.

When compared by means of t-tests for independent samples, none of the

Table 3.29: Comparison of choosers and non-choosers in SEC-4: Test scores.

	Choosers	Non-choosers
Cloze	64 (5.35)	62 (4.50)
PHO-LI	31 (2.24)	31 (2.48)
PHO-RE	30 (3.50)	29 (2.97)
LEX-LI	31 (2.77)	29 (3.49)
LEX-RE	33 (2.95)	33 (2.53)
MS-LI	26 (3.33)	24 (3.93)
MS-RE	28 (3.04)	26 (3.54)

seven tests were significantly different. Only the differences for the LEX-LI and both MS tests were marginally significant - to the advantage of the choosers, as can be observed in Table 3.29. In other words, the choosers are only slightly better learners of French than the non-choosers, at least according to the measures we used.

Table 3.30: Comparison of choosers and non-choosers in SEC-4: Questionnaire data.

	Choosers	Non-choosers
Languages, lower grades	7.5 (0.61)	7.3 (0.66)
French, lower grades	7.9 (0.73)	7.4 (0.71)
Languages, higher grades	7.0 (0.59)	6.4 (0.51)
French, higher grades	7.3 (0.79)	6.3 (0.66)
French lessons	3.7 (0.61)	2.8 (0.66)
French language	4.2 (0.52)	3.5 (0.65)
Future utility	2.4 (0.87)	0.9 (1.00)

More and larger differences, however, were found in the questionnaire data shown in Table 3.30. The school marks in the higher grades of secondary school, both for French and for all modern languages taken together, were significantly higher for the choosers. Also, all three attitudinal questions yielded significantly higher scores for the choosers.

In other words, although their actual test performance was only slightly better, the choosers' attitude was much more positive, possibly as a result of their superior school performance in French. In addition, they performed better on modern languages in general in the higher grades of secondary school. Thus, there is some indication indeed that those who are more talented language learners tend to choose French more often than those who are not. Interestingly enough, though, the average number of foreign languages chosen by the non-choosers of French was 1.4. Taking into account that one foreign language is compulsory in Dutch secondary schools, this means that 40% of the students still voluntarily took a second foreign language - in this case German (cf. section 2.1). On the other hand, the situation was not such that all choosers of French by definition took *three* modern foreign languages (English, French, German): since the average number in this group was 2.32, only 32% also took German, apart from the compulsory language, English, and the second one, French, on account of which they had been included in our sample.

3.12. THE ROLE OF ATTITUDES, MOTIVATION AND USE

Although our questionnaire did not really contain an attitude and/or motivation scale, there were three questions that probed aspects of attitudes and motivation (cf. section 2.5.7): the opinions on French lessons in school, on French as a language, and on the potential future utility of French. Since the correlations between the three items were reasonable (Lessons x Language: .28; Lessons x Utility: .31; Language x Utility: .39), we decided to treat them as a scale: we constructed an attitude/motivation index (AMI) by summing the scores on the three items. The scores of our six groups of subjects on this AMI are given in Table 3.31. It appears that there were marked differences between the two education levels - as might have been expected on the basis of the data discussed in section 3.11 - but the scores remained relatively stable over time.

The questionnaire also contained a question about out-of-school contact with the target language - the question we used to eliminate subjects who had had frequent and intensive contacts with French (cf. section 2.6). The number of domains - *holiday*, *friends and relatives*, *literature*, *media*, and *otherwise* - they ticked in this question, were counted: that means the score could theoretically run from 0 to 5. We will be referring to this score as

Table 3.31: Group means (s.d.) on the attitude/motivation index.

		Years of non-use		
		0	2	4
Years of training	6	10.1 (1.73)	9.2 (1.81)	9.2 (1.68)
	4	7.2 (1.18)	6.6 (1.52)	8.3 (1.49)

'contact'. The group means are given in Table 3.32.

Table 3.32: Group means (s.d.) on the contact variable.

		Years of non-use		
		0	2	4
Years of training	6	1.6 (1.04)	1.4 (1.08)	1.4 (0.87)
	4	0.6 (0.95)	0.4 (0.58)	1.0 (0.89)

Robert Gardner has repeatedly stressed the important role that attitutes and motivation would play in language attrition. In his initial article on this topic (Gardner 1982), he speculated: "Attitudinal/motivational variables could also influence second language retention by orienting the individual to take every opportunity to maintain proficiency in the language" (p. 32). In subsequent investigations (Gardner et al. 1985, 1987) the hypothesis was confirmed indeed, and the point was taken by other authors too (Edwards 1976; Snow et al. 1984 - cf. section 1.1).

If the hypothesis applies in our case as well, one would expect the following effects to occur:

(1) a substantial correlation between AMI and contact scores;
(2) substantial correlations between AMI scores and all test scores;
(3) substantial (negative) correlations between AMI and attrition scores.

The first effect did occur indeed: the correlation between AMI and contact

was .44 (n=150; p<.001), but it was of course flattered by the fact that the subjects with 6 years of French training had both higher AMI and higher contact scores. Therefore, the correlation was computed for the two education levels separately, with the result that it was lower in either case: .25 for Education-6 (n=75; p=.016) and .34 for Education-4 (n=75; p=.001). None the less, these are quite respectable correlations if we keep in mind that all our subjects were in fact people with infrequent contacts with the target language: the mean score on the contact variable was 1.1, meaning that on average the subjects ticked only slightly more than one domain, which, incidentally, was *holiday* in almost 50% of the cases.

The second effect, the correlations between AMI scores and test scores, appeared to be present at first sight: for the sample as a whole the correlations ranged from .19 to .53. However, when we looked at the same correlations at each of the education levels, they were considerably lower. For Education-6 they ranged from -.11 to .31, but the majority was below .20; for Education-4 they ranged from -.04 to .45, but the preponderance fell below .30. So, across the board, the relationship was a rather weak one.

Concerning the third effect, the relationship between AMI and amount of attrition, we can be brief: it simply did not exist. The only test scores that we could use in this case were, of course, those of the morpho-syntactic tests - since they were the only ones exhibiting attrition - for groups D and E - since they were dependent measurements. For the MS-LI test the correlation was -.02 (n=25; p=.461); for the MS-RE test it was .16 (n=25; p=.218).

In summary, then, we might say that there was only a weak relationship between attitudinal/motivational factors and language retention in our case, but we have to add in all fairness that our instruments may not have been the most adequate for determining such a relationship. Another conclusion that follows from the data discussed here is that, as Bahrick (1984) also noted, substantial amounts of knowledge are retained even without a lot of intermediate exposure to the target language.

4. CONCLUSIONS AND DISCUSSION

In this chapter we will draw the conclusions from our exploratory investigation of the long-term retention - across two and four years - of French foreign-language skills acquired in a training period of either four or six years in Dutch secondary education. The design of the investigation was such that we compared six groups of 25 subjects each on a number of receptive language measures that tested general receptive proficiency; listening and reading comprehension; and the mastery of certain phonological, lexical, and morpho-syntactic properties of French, in particular those aspects that contrasted with Dutch. In addition, a number of self-evaluation measures were administered.

In the sections to follow we will first of all evaluate the methodology employed (4.1). Subsequently, our findings with respect to the *quantity* of attrition are discussed in terms of the different potential forgetting curves discussed in chapter one (4.2). The findings with respect to the *quality* of attrition are summarized in the next section (4.3), where they will be evaluated in the light of different cognitive learning theories. The last section (4.4) contains some suggestions for future research.

4.1. METHODOLOGY

As far as the methodology of the investigation is concerned, quite a number of aspects deserve further evaluation at this point. In the subsequent sections we will discuss the overall design (4.1.1), the instruments (4.1.2), and the subjects (4.1.3).

4.1.1. The overall design

In section 2.3 we explained how we arrived at a design which involved a mix of longitudinal and cross-sectional measurements. Although this mix implied a sacrifice in our analyses of the data - in the sense that we were unable to analyze all six groups in one analysis of variance; cf. chapter 3 - the so-called bootstrap approach (see Schils 1988) ensured that we could still answer all the questions we had intended to. In other words, a *statistical* solution to the analytical problem was found. Nevertheless, we should

point out at the same time that the *methodological* problems inherent in mixing both types of measurements remained, of course. In chapter 3 we argued that the best analysis was a straightforward ANOVA of the 'cornerstones' of our design, i.e. the four groups with 0 and 4 years of non-use. Even then we had to acknowledge that there was a potential threat from the 'testing' factor, because one of those four groups was tested for the second time, while the other three were tested for the first time (see Table 2.1 for more details).

When the research plan for the project (Van Els & Weltens 1983) was written, we built in a check for possible cohort effects: the groups that would be tested longitudinally in the second test round, were also tested cross-sectionally in the first test round. This check proved ineffective. As explained in section 3.1, the data suggested the presence of a cohort effect, certainly in the case of the non-choosers (cf. Table 3.2): there seemed to be differences between groups in terms of school marks for French. However, when we carried out analyses of covariance in which we controlled this factor, the results hardly differed from the straightforward analyses of variance. In other words, the check we had built in suggested the presence of a cohort effect, while the data from the actual groups concerned denied it.

The design was chosen on the basis of the assumption that number of years of training - or, rather, proficiency level attained - and number of years of non-use would be critical factors in the situation we wanted to investigate. We chose two obvious proficiency levels - obvious, that is, in terms of Dutch secondary education; cf. section 2.1 - namely six and four years of training; similarly, we used the interval of two years in choosing the levels of our second factor, non-use. Since we wanted to have three points of measurement, the obvious choice was an interval of two and four years following the acquisition period.

We were aware, of course, of the important role that post-course exposure to the foreign language in question would play in the attrition process. That means we basically had two options: either to eliminate the variable, namely by selecting subjects that did not vary in terms of post-course contact, or to measure post-course contact and incorporate it into our design. The latter option was chosen in the studies by Bahrick (1984) and Gardner et al. (1985, 1987).

Bahrick (op.cit) had his subjects indicate how often they had spoken,

listened to, read, or written Spanish since their last Spanish course. He gave them a four-point scale marked (1) "once a year or less"; (2) "2-11 times per year"; (3) "several times each month"; (4) "several times each week". The listening scale was broken down into radio, television, films, and conversation *of others*; the reading scale elicited separate ratings for newspapers, magazines, books, and correspondence. In addition, subjects were asked to estimate the average duration of each type of activity. The total 'exposure time' was calculated for each activity, and entered into the analysis. Unfortunately, most subjects did not find - or seek? - many opportunities to use Spanish, with the result that the analysis did not produce any relevant effects. However, one may well wonder how accurate these data were, especially since Bahrick's subjects had to look back over very long periods of time, up to no less than fifty years. Moreover, it remains unclear whether this approach of 'total exposure time' is a valid indicator of subjects' contact with the foreign language. For example, it fails to take into account the *quality* of the exposure (Was the t.v. just switched on, or did the subject actually watch the programme?), nor its *spacing in time* (Did the contact occur at regular intervals or was it concentrated in one particular smaller period of time?). With respect to the latter point, the psychological literature on the so-called 'spacing effect' is, of course, relevant. Rea & Modigliani (1988:403), for example, conclude that "the few studies that have investigated the spacing effect in applied settings ... have shown that distributed practice leads to significant better retention than massed practice".

Gardner et al. (1985) used an approach similar to Bahrick's, but they were at an advantage compared to Bahrick in the sense that they only covered the period of the summer vacation, which increases the likelihood that subjects were able to provide valid information. However, in a later study (Gardner et al. 1987), a more global and, therefore, maybe more realistic approach was adopted. The subjects were asked to indicate on a five-point scale ranging from (1) "not at all" to (5) "very frequently" how often they had used the language - in this case, French - in nine different contexts, e.g. "speaking French to friends", over a period of three months.

There are more examples of (preliminary) attempts at capturing the contact variable, mainly in language acquisition studies (e.g. Day 1985; Lightbown & Spada 1987), where it is 'measured' because the investigators want to know how much out-of-school contact the subjects have with the foreign language *while they are still learning it in a classroom context.* But

these attempts can add very little to the two methods discussed above, which were used in a language attrition context.

In all, we think the quantification of (post-course) exposure is still problematic. Therefore, we are at the moment carrying out a pilot study into the validity of different retrospective self-report measures of exposure, such as the ones above (to be reported in Achterberg 1988). Unfortunately, though, the results of this work could not be used in the present investigation.

In fact, we chose the other option available in terms of controlling the influence of the contact variable: we eliminated it by selecting subjects that had had hardly any contact with the target language outside the school context. The fact that very few subjects failed to meet this criterion proves that this choice was not only practical, but also ecologically valid: in our case, once formal training has stopped, having hardly any contact appears to be the rule, rather than the exception.

4.1.2. The instruments

In section 2.5 we have paid ample attention to the relatively low reliability of our (self-developed) tests, but we also hope to have shown that there were good reasons for using them anyway. One of the arguments given there for the assumption that our tests were more reliable than the reliability coefficients in the pilot tests suggested, was that we used a relatively small and homogeneous sample in our pilot tests. Now, in hindsight, we are in a position to compute the reliablility of the tests on the basis of the actual sample measured in the two tests rounds, consisting of 150 subjects of much more varied proficiency levels. These reliability coefficients are presented in Table 4.1, along with the coefficients computed across the so-called composite scores of subsets of items (Alpha'; see section 2.5).

The results showed that the reliability as determined across the actual test sample was indeed higher in most cases. Interestingly enough, all values approached the 'composite score alpha' (Alpha') quite closely, which confirmed our assumption that the latter, although based on computations in which the number of 'items' was relatively small (cf. Table 4.1), was a much better estimate of the actual test reliability than the straightforward alpha.

Our integrative tests - i.e. the test of general receptive proficiency, the listening comprehension test, and the reading comprehension test - should

Table 4.1: Reliability (Cronbach's alpha) of the self-developed tests in the pilot test (n=18) and in the actual test sessions (n=150).

	Pilot test: Alpha	Pilot test: Alpha'	Actual test: Alpha
PHO-LI	.44 (n=34)	.69 (n=8)	.62 (n=34)
PHO-RE	.75 (n=34)		.75 (n=34)
LEX-LI	.68 (n=40)	.77 (n=8)	.75 (n=40)
LEX-RE	.27 (n=40)		.70 (n=40)
MS-LI	.72 (n=40)	.80 (n=4)	.80 (n=40)
MS-RE	.50 (n=40)		.78 (n=40)

be treated separately here. They had been pretested and revised more profoundly, in the first case by ourselves, in the second and third cases by CITO (cf. sections 2.5.1 through 2.5.3). In addition, their main problem is not their reliability, but their validity. We have already hinted at the growing scepticism about the validity of the cloze test as a measure of general language proficiency, and some people had particular reservations about us using it in a multiple-choice format, especially because of its supposedly lower reliability. As far as the latter point is concerned, however, we hope to have shown that the multiple-choice version, apart from having somewhat less discriminatory power than the open-ended version, was in fact highly comparable. Even its reliability was in the same range: as measured across the 100 subjects from our sample that completed both versions (see section 3.2), the open-ended version had a reliability of .90; the MC version reached .86, confirming Hinofotis & Snow's (1980) conclusion that the conversion of an open-ended cloze test into an MC format reduces its reliability only marginally.

In connection with the particular reading comprehension test that we also used, Maas-de Brouwer (1983:183) remarked that "adults score significantly higher than younger candidates" (our translation). However, this finding is not surprising in view of the fact that these tests are validated for adolescent exam candidates. Apparently, intellectual maturation influences the scores on tests such as these, which makes them less well suited for investigations such as the present one.

We could add here that several authors have specifically warned against the use of (objective) integrative tests for measuring language attrition.

Jaspaert & Kroon (1987) argue that metalinguistic skills and testpertise are imminent dangers for especially such tests, and stipulate that the tests used should "contain the linguistic elements that are most susceptible to loss" (o.c.:92). One reservation that should be made here is that they are referring to the investigation of L1 loss, where a number of aspects of the situation may demand a particular methodology; for one thing, the proficiency to be measured is undoubtedly higher than in any investigation of FL attrition.

Gardner et al. (1987:39) reported "only slight evidence for language attrition in the objective assessments of proficiency and even some apparent suggestion of language growth", and concluded: "rather than indicating little language loss or even growth, it is probably more likely that these results underscore the difficulty of measuring change using objective measures" (l.c).

One cannot help but wonder, though, why such tests would be suspect - either *a priori*, as in the case of Jaspaert & Kroon (1987), or *a posteriori*, as in the case of Gardner et al. (1987). The argument that integrative tests may not be the most sensitive in terms of detecting early signs of attrition seems to us not be a conclusive one. Surely, when they adequately measure a relevant skill, the results have to be taken seriously. In other words, until proven otherwise, Gardner et al.'s (1987) results mean *to us* that their subjects' proficiency remained unchanged or even improved in some areas, and decreased somewhat in others. For, to use Jaspaert & Kroon's (1987:92) words, "a certain influence of especially metalinguistic knowledge may not be unacceptable or even undesirable". The fact that subjects' self-reports appear to be far more sensitive in detecting signs of attrition (cf. Gardner et al. 1987:37-38) is, in our view, no reason whatsoever to doubt the validity of certain test results. It is still an open question whether self-reports accurately reflect (loss in) language proficiency; it may well be that they only correlate highly with more sensitive language proficiency measures, such as reaction times in lexical-decision experiments, or that they should even be treated as separate data altogether, indicating how subjects perceive their own competence, rather than how they think they would perform on actual tests. We will return to this point in section 4.3 below, when evaluating our own results.

An entirely different point in connection with global measures is raised by Moorcroft & Gardner (1987): global measures can hardly provide information as to the *types of skills lost*, and it is exactly this information that

might enable "educators and researchers (...) to develop a better understanding of processes operating in the acquisition of the language that they might strengthen to promote second-language retention" (o.c.:328).

Regarding the phonological, lexical, and morpho-syntactic tests we developed, a number of evaluative comments are also in order. The phonological tests proved to be too easy for our subjects: a clear ceiling effect occurred in three out of four sub-categories of items, with only the fourth (contrast-consonantal) leaving any room for improvement. Similarly, one could argue that three of the four sub-categories in the lexical tests were on the easy side. These ceiling effects may have been one of the reasons why we found relatively little attrition (cf. section 4.2 below), in the sense that these tests may not have had enough discriminatory power.

A particular problem in connection with the lexical tests was the definition of 'cognates'. We widened our definition compared to the ones that some other researchers have used, but we feel that the resulting list of targets (cf. Appendices 6 and 7) was intuitively satisfactory, at least in the sense that the cognates were sufficiently different from the non-cognates. In point of fact, this is a criterion that deserves serious consideration: (systematically elicited) perception of 'sameness' might well be the best criterion.

With respect to the morpho-syntactic tests, we should point out that the two items dealing with the use of the subjunctive should not have been included in the test: their correct scores were far too low, even for choosers of French in SEC-6. This means the phenomenon is not really learned, as was confirmed by teachers' reports collected by Grendel & Poppe (1986).

4.1.3. The subjects

Several aspects of our selection procedure proved rather problematic. In practice, we had to allow more variability in terms of background variables than we had anticipated. This was a direct result of the fact that we tried to control as many background variables as we could, rather than just the ones known to have an influence on the attrition process. Unfortunately, however, such knowledge is as yet unavailable.

A second problem might lie in the fact that our secondary school subjects formed - apart from a few eliminated subjects - complete classes, whereas the other subjects were university students, and as such really

represent a sub-group of all ex-students of secondary school, namely the ones that enter this specific type of tertiary education. Again, this may have been one of the reasons why we have ascertained relatively little attrition.

In addition, as noted above, global measures of proficiency seem to be relatively susceptible to maturation effects. Since the factor non-use ran parallel with the maturation of our subjects between the ages of 18 and 22 in the case of the non-choosers, and the ages of 16 and 20 in the case of the non-choosers, another depressing effect on the amount of attrition may have occurred here. This is all the more likely in view of the fact that our subjects have continued reading in one or more foreign languages other than French during their studies.

4.2. THE RATE OF ATTRITION

Although our results revealed substantial attrition only in the self-assessments and the morpho-syntactic tests, we are still in a position to draw conclusions on the pattern of language attrition, simply because these results all point in the same direction. They quite clearly indicate that attrition sets in rather quickly, and then levels off. This is in line with the traditional theory of forgetting as formulated by Ebbinghaus (1885), and it confirms what Kennedy (1932), Flaugher & Spencer (1967), Godsall-Myers (1981), and Bahrick (1984) have reported.

Secondly, the absence of interactions between training level and non-use in the cases where attrition was found indicates that attrition is - in absolute terms - independent of training level. In other words, subjects lose a fixed amount of knowledge, independent of their original level. In relative terms, however, this means that high-proficiency subjects lose less (cf. Fig. 1.3 IIa and IIb). This finding is *not* in line with the traditional theory on forgetting, which states that a certain *proportion* of one's knowledge is lost irrespective of the total knowledge (cf. Fig. 1.3 Ia and Ib). However, as noted in section 1.2, findings similar to ours have been reported by Smythe et al. (1973), Godsall-Myers (1981), and Bahrick (1984). We would, therefore, like to quote once more Bahrick's (1984:116) conclusion on this point:

> "the total *amount* to be forgotten ... is relatively constant for individuals at different levels of training, but this amount becomes a progressively smaller portion of total knowledge with higher levels of training".

At the beginning of this paragraph we noted that substantial attrition was only found for the morpho-syntactic tests. Strictly speaking, we could have added the written lexical test, since the effect of non-use was significant there. But when we look at the mean scores, all the loss amounts to is one or two lexical items across a time interval of four years; in relative terms, a loss of 3 to 6 per cent of the original knowledge. This may be *statistically* significant, but one can hardly argue that it represents a loss that is *communicatively* significant; or, to use Andersen's (1982:85) terminology, whether this loss is "dysfunctional", or just "cosmetic" (cf. also Van Els & Weltens 1987).

On the other hand, we have the results of our pilot study (reported in Verkaik & Van der Wijst 1986) and the self-reports: both clearly indicated loss of vocabulary after two years of non-use already. An explanation for the discrepancy between the results of our pilot study and the results of the actual investigation may be that the task used in our pilot study - the lexical-decision paradigm - is a much more sensitive one. For one thing, the lexical test in the actual investigation was - like all the other tests in fact - self-paced, i.e. subjects had ample time to squeeze out of their memories anything that was still there, however vaguely and remotely. In a lexical-decision task, the *speed* of the access to lexical items in the mental lexicon is measured. It is quite conceivable, therefore, that time pressure would have revealed larger differences also in our lexical tests. Maybe this is also why the self-reports show much heavier attrition: subjects may not only have the impression that their vocabulary knowledge has suffered from non-use, they may be absolutely right, in the sense that the accuracy may not have suffered (yet), but that the speed has. (We will return to this point in the next section, where we will be dealing with the question what is lost.) Incidentally, this same argument may also apply to other tests used, especially to our test of general (receptive) proficiency, the multiple-choice cloze test, if only because it partly relies on lexical skills.

For the morpho-syntactic level, the situation is much clearer. Here, both self-reports and test results tell us the same thing: there is heavy attrition, already after two years, amounting to 9 to 14 per cent of the original knowledge. The self-reports are, again, more dramatic: they go down from 3.9 to 2.2 (choosers) and from 3.2 to 1.9 (non-choosers) on a five-point scale. In terms of the anchor points provided, they go down from "reasonable"/"good" to "bad". As noted before, in absolute terms, the self-reports for grammar exhibited the steepest fall. An interesting aspect of the

self-reports in relation to the test scores is that the *rank order* of the three linguistic levels is the same in both cases: phonology increased in the tests and decreased only weakly in the self-reports; vocabulary stayed about the same in the test results and decreased more strongly according to the self-reports; finally, grammar decreased in the tests and showed the largest decrease in the self-reports.

The overall picture presented here with respect to the three linguistic levels tested - phonology, vocabulary, grammar - was confirmed in a factor analysis of all tests scores. A two-factor solution emerged with the first factor, 'vocabulary plus grammar', showing a decrease over time for both training levels, and the second factor, 'phonology', increasing over time for both training levels.

4.3. THE NATURE OF ATTRITION

We have already indicated in the previous section that the three linguistic levels investigated were affected quite differentially by non-use - although, strictly speaking, the tests cannot be compared in a straightforward manner, of course. What can be done, however, is drawing conclusions about the different categories of items distinguished in each of the tests. Before doing so, however, one preliminary point should be made. In all cases, as we have seen in sections 3.5 through 3.7, the development in time - whether positive or negative - consisted of gain of some elements, retention of others, and loss of yet others. An important consequence of this fact was that, even in the case of significant gain, we could identify areas that showed systematic attrition, namely in the case of the phonological tests (see section 3.5).

On all three linguistic levels, then, one of Andersen's (1982) *linguistic feature hypotheses* was confirmed: in phonology, vocabulary and grammar, *contrast between the native and the target language* appeared to play an important role. In the written phonology test, the category of contrasting vocalic oppositions, as in *en-on*, *tendu-tondu* and *répand-répond*, was the only category suffering attrition. In the written lexical test, the analysis of variance indicated that non-cognates are more susceptible to attrition than cognates; an analysis on the level of the individual items revealed that non-cognates of low-frequency were in particular subject to attrition. Similar results were found in our lexical-decision experiment (reported in Verkaik &

Van der Wijst 1986): reaction times to target words highly comparable to the ones in our lexical tests had increased significantly after two years of non-use, but the increase - i.e. the deceleration in the lexical access - was about 50 ms. for all high-frequency words and for low-frequency cognates, but it amounted to about 80 ms. for low-frequency non-cognates. Finally, the grammar tests indicated that contrasting grammatical phenomena are more difficult to retain than phenomena that the target language shares with the native language; on the item level, this could be further specified in the sense that the largest problems were found in the contrasting aspects of the pronominal system, again confirming what we had found in a pilot study (Grendel & Poppe 1986; Weltens et al. 1986b). On the other hand, it might be noted that, although 'contrast' in general played its expected role, 'degree of contrast' did not, at least not as might have been expected on the basis of the interference theory: attrition did *not* increase with the subtlety of the contrasts - although it is unclear whether the subdivision we used is psycholinguistically real.

Finally, the most important thing of all to be noted here in connection with the question of "What is lost?" is that general receptive proficiency in French is clearly not subject to attrition after four years of non-use, whereas grammar - and to some degree also vocabulary - clearly is.

A recent study by Moorcroft & Gardner (1987), which investigated the attrition of oral production in French over a period of three months, revealed significant declines for different global measures of oral production. However, with respect to the role of vocabulary and grammar in this process they concluded:

> "The observed decline in quantity (both in terms of grammatical units produced, clauses [judged to be successful communications], and in terms of speaking time) may therefore not be related to the fact that subjects have less vocabulary available with which to construct sentences, but to the fact that they have more problems in using grammatical rules to link vocabulary in sentences" (o.c.:333).

It should be noted, however, that - as in our case - vocabulary was tested by means of a task that was self-paced. No time pressure was exerted, nor did they measure the amount of time required to produce the items. As already noted in section 4.2 above, our point is that the subjects may not have differed in the number of vocabulary items they could actively recall,

but that they may have differed in the *speed* with which they could do so. In other words, the fact that the oral production process was slowed down, certainly was the result of problems with applying grammatical rules, but it may well have been partly due to problems with recalling lexical items.

Another important conclusion drawn by Moorcroft & Gardner (1987:339) is that, when looking at the grammatical losses in more detail, it appears that "a thoroughly learned structure is relatively immune to language loss". In other words, Moorcroft & Gardner (o.c.) assume that at the level of individual items - be it phonemic oppositions, or lexical items, or grammatical rules - there is a certain level of mastery that, to borrow Neisser's (1984) phrase, "confers immunity against forgetting". In connection with self-report results Gardner et al. (1985) made a similar point: for speaking and understanding, "There was not any significant loss on skills that had been overlearned or on skills for which the original level of competence was low. Instead, students demonstrated attrition on medium-level skills" (p. 529).

When we extend this argument from the *within*-subjects level to the *between*-subjects level, one might argue that subjects that attain a certain (high) degree of mastery of the target language are, as a result, less prone to lose their language skills. This, then, echoes the hypothesis put forward by Pan & Berko-Gleason (1986:204) that there might be a "critical mass of language that, once acquired, makes loss unlikely", and Neisser's (1984:33) conjecture that there might be a "critical threshold during learning" beyond which isolated responses, or 'facts', become part of "mental representations of complex information structures" with the result that they - at least for some time - "become immune to interference or decay". We already discussed this point in section 1.3, when trying to explain why a number of studies found no loss, or no initial loss at least (see also Fig. 1.2), and there we concluded indeed that there are indications that high-proficiency subjects seem to retain most of their language skills for at least some time. If this is the case indeed, one might argue that the subjects in the present study have acquired receptive skills in French to a degree over and above this "critical threshold" - after all, they had had no less than 400 to 600 hours of French in four and six years respectively - with the result that, with the exception of some grammatical knowledge and the speed of vocabulary retrieval, their skills were relatively immune to attrition, at least during a period of four years.

The fact that we failed to find any attrition on the more general levels of

foreign-language proficiency, although there are clear indications of attrition on the levels of vocabulary and grammar (cf. also Weltens & Van Els 1986), has an interesting parallel in quite a different research tradition, namely 'memory for text'. In this research area it is generally assumed that readers build up a hierarchical stucture of the text while reading it, or even fill in slots in an existing 'text grammar', a so-called 'schema'. When tested for recall, the so-called 'levels effect' invariably shows up: subjects recall higher proportions of high-level information, i.e. information that is on a higher level in the text structure (see, for example, Thorndyke 1977). Recently, this paradigm has been applied by Naveh-Benjamin (1988) to the information taught in an entire university course on methodology and statistics. The subjects who took the course in question were retested one and two years after completion of the course. Some forgetting occurred after one year already, but "primarily in the lower levels of the structure" (p. 383). If these findings also apply to language learning, they could explain the result - at first sight somewhat contradictory - that 'lower-level' skills decrease somewhat, without 'higher-level' skills suffering as well.

A second recent study into the role of schemata is Bahrick et al. (1988). However, this study operationalized the schema in quite a different way: it investigated whether familiarity with Spanish word-formation rules would facilitate the learning and retention (across two weeks) of Spanish neologisms. Bahrick et al. (1988:388) concluded that "a general schema for the Spanish language does not play a significant role in facilitating the learning and retention of new vocabulary".

Another - related - theory that may be relevant in terms of explaining our results, is Ausubel's (1963, 1968) *subsumption theory*, in which *meaningful* learning - as opposed to *rote* learning - is characterized as relating new pieces of information to already existing cognitive structures, "hanging new items on existing cognitive pegs", as Brown (1987:65) described it. Brown (o.c.:67) also notes that the basic idea behind the theory is at least a century old; he quotes the following passage from William Jones, which dates back to 1890:

> "In mental terms, the more other facts a fact is associated with in the mind, the better possession of it our memory retains. Each of its associates becomes a hook to which it hangs, a means to fish it up by when sunk beneath the surface. Together, they form a network of attachments by which it is woven into the entire tissue of our thought".

Brown adds that interference phenomena "have relatively little influence on meaningful material, and retention is highly efficient" (p. 68). He then goes on to describe forgetting of meaningful material as a "second or 'obliterative' stage of subsumption": "the specific items become progressively less identifiable as entities in their own right until they are finally no longer available and said to be forgotten" (l.c.) - in fact, very much alike the processes assumed to be operative in the levels effect - which would explain why meaningful material exhibits "retention beyond that normally expected under more traditional theories of forgetting" (o.c.:69). In other words, an argument along the lines presented in connection with the "critical mass" or "critical threshold" above - our subjects had acquired French over and above such a level - could be given here: assuming that for all our subjects learning French had been "meaningful learning", very little attrition would be expected indeed.

4.4. DIRECTIONS FOR FUTURE RESEARCH

In chapter one we argued that attrition research may ultimately contribute to a better understanding of the learning process, and to *promoting*, rather than just *measuring*, retention. Since this was essentially an exploratory project, one should not expect too much from it in this respect. Nevertheless, a few general observations ought to be made. For one thing, the general picture we found was a reassuring one. Whatever qualifications we will make below, one cannot deny that our subjects' retention of their school-learned French was impressive. The possible explanations we offered suggest that it may have been their high level of proficiency that ensured this high retention rate, but further research dealing directly with this question is required. By the way, the high retention rate was observed for *both* training levels. In other words, contrary to what some people seem to think, *four* years of French training in Dutch secondary education also leads to a level of proficiency that lasts.

On the other hand, we tested only *receptive* skills. One can only surmise what the results would have looked like, had we also administered productive measures. In addition, some of the measures we *did* use may have lacked the discriminatory power to detect (beginning) attrition. In general, one might conclude that measures involving a certain degree of time pressure are superior in this respect to self-paced, or leisurely paced tests. Again, however, further research is needed on this point.

Even given the fact that only receptive skills were tested, and given the limited qualities of at least some of our measures, we found clear indications of attrition in the area of grammar, more specifically in the (contrastive domains of the) pronominal system, and probably also (one area of) vocabulary, namely in the case of non-cognates of low frequency. And we are also left with systematic self-reports of attrition in *all* areas questioned. The least we can conclude from this is that, if we assume that actual attrition is moderate indeed, the backgrounds for these pessimistic self-reports warrant further investigation. It may have been the case that, although the self-report questions were clearly designed to elicit problems with *receptive* skills, subjects did not manage to rule out the problems they had encountered in using French *productively*, when completing the questionnaire. A possibility that should not be overlooked either is that the self-reports are better indicators of our subjects' competence than (some of) the tests used.

Finally, we might conclude that a period of non-use of four years simply is too short to find heavy attrition with the subjects in question. Alternatively, further research ought to address the question to which degree the results reported here can be generalized to subjects who have received French training with teaching methods that are fundamentally different from the ones currently employed in Dutch secondary schools.

REFERENCES

Achterberg, A. (1988), Het meten van de mate van contact met een vreemde taal. Master's thesis University of Nijmegen.

Aertssen, J., M. van de Berg, R. Buursen & T. Claassen (1985), Tweede-taalverlies: Verlies van het Engels bij Nederlandse kinderen. Unpublished MS Department of Applied Linguistics, University of Nijmegen.

Allendorff, S. (1980), *Wiedererwerb einer Zweitsprache, dargestellt am Beispiel der englischen Negation.* Kiel: Department of English, University of Kiel (*Arbeitspapiere zum Spracherwerb* 25).

Andersen, R. (1982), Determining the linguistic attributes of language attrition. In: R. Lambert & B. Freed (eds.), 83-118.

Appel, R. (1983), Buitenlanders en taalbehoud. *Interdisciplinair Tijdschrift voor Taal- en Tekstwetenschap* 3:152-166.

Ausubel, D. (1963), Cognitive structure and the facilitation of meaningful learning. *Journal of Teacher Education* 14:217-221.

Ausubel, D. (1968), *Educational psychology: A cognitive view.* New York: Holt, Rinehart & Winston.

Bahrick, H. (1984), Fifty years of second language attrition: Implications for programmatic research. *Modern Language Journal* 68:105-118.

Bahrick, H., C. Wellman & L. Hall (1988), The effect of language schemas on learning and retention of vocabulary. In: M. Gruneberg et al. (eds.), 385-389.

Baltes, P., H. Reese & J. Nesselroade (1977), *Life-span developmental psychology: Introduction to research methods.* Monterey, CA: Brooks/Cole.

Bauman, J. (1980), *A guide to issues in Indian language retention.* Washington, DC: Center for Applied Linguistics.

Berko-Gleason, J. (1982), Insights from child language acquisition for second language loss. In: R. Lambert & B. Freed (eds.), 13-23.

Berman, R. & E. Olshtain (1983), Features of first language transfer in second language attrition. *Applied Linguistics* 4:222-234.

Bijlsma, J. (1976), Onder de loep 1: Een vergelijkende analyse van de meest gebruikte leergangen Frans. Unpublished MS Department of French, University of Nijmegen.

Boyd, S. (1986), Using the present to predict the future in language contact: The case of immigrant minority languages in Sweden. In: B. Weltens et al. (eds.), 99-115.

Brewer-Bomar, K. (1981), Second language lexical and syntactical interference on the first language of two four year old Spanish speakers. Diss. University of Pittsburgh (*Dissertation Abstracts International* 42:5105-A).

Brown, H. (1987), *Principles of language learning and language teaching.* Englewood Cliffs, NJ: Prentice-Hall (2nd ed.).

Burling, R. (1978), Language development of a Garo and English-speaking child. In: E. Hatch (ed.), *Second language acquisition.* Rowley, Mass.: Newbury House, 54-75.

Butcher, A. (1976), *The influence of the native language on the perception of vowel*

quality. Kiel: Institute of Phonetics, University of Kiel (*Arbeitsberichte* 6).
Campbell, C. (1980), Second generation Australian Italian in Giru, North Queensland. *ITL Review of Applied Linguistics* 49-50:5-22.
Campbell, D. & J. Stanley (1966), *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
Campbell, R. & J. Schumann (1981), Hypnotism as a tool in second language research. In: R. Andersen (ed.), *New dimensions in second language acquisition research*. Rowley, Mass.: Newbury House, 80-91.
Capelle, J. & G. Capelle (1969), *La France en direct*. Paris: Hachette.
Caramazza, A. & E. Zurif (eds.) (1978), *Language acquisition and language breakdown. Parallels and divergencies*. Baltimore, MD: The Johns Hopkins University Press.
Chatterjee, S. (1984), Variance estimation in factor analysis: An application of the bootstrap. *British Journal of Mathematical and Statistical Psychology* 37:252-262.
CITO (1983), *Leerdoelgerichte toetsen Frans luistervaardigheid*. Arnhem: National Institute for Educational Measurement.
Claessen, J., A. van Galen & M. Oud-de Glas (1978), *De behoeften aan moderne vreemde talen. Een onderzoek onder leerlingen, oud-leerlingen en scholen*. Nijmegen: Institute for Applied Sociology (*Studies over het onderwijs in de moderne vreemde talen* deel IV).
Clark, E. & B. Hecht (1983), Comprehension, production, and language acquisition. *Annual Review of Psychology* 34:325-349.
Clark, J. (1981), Language. In: T. Barrows (ed.), *College students' knowledge and beliefs: A survey of global understanding*. New Rochelle, NY: Change Magazine Press: 25-35 and 87-100.
Clark, J. (1982), Measurement considerations in language attrition research. In: R. Lambert & B. Freed (eds.), 138-152.
Clark, J. & E. Jorden (1984), *A study of language attrition in former U.S. students of Japanese and implications for design of curriculum and teaching materials. Final project report*. Washington, DC: Center for Applied Linguistics (*ED* 243 317).
Clyne, M. (1980), Typology and grammatical convergence among related languages in contact. *ITL Review of Applied Linguistics* 49-50:23-36.
Clyne, M. (1981), Second language attrition and first language reversion among elderly bilinguals in Australia. In: W. Meid & K. Heller (eds.), *Sprachkontakt als Ursache von Veränderungen der Sprach- und Bewusstseinsstruktur. Eine Sammlung von Studien zur sprachlichen Interferenz*. Innsbruck: Institute for Linguistics, 25-32.
Cofer, C. (1984), Comments on "Semantic memory content in permastore: Fifty years of memory for Spanish learned in school" by Bahrick. *Journal of Experimental Psychology: General* 113:30-31.
Cohen, A. (1974), Culver City Spanish immersion program: How does summer recess affect Spanish speaking ability? *Language Learning* 24:55-68.
Cohen, A. (1975), Forgetting a second language. *Language Learning* 25:127-138.
Cohen, A. (1986), Forgetting foreign-language vocabulary. In: B. Weltens et al. (eds.), 143-158.
Craik, F. & R. Lockhart (1972), Levels of processing: A framework for memory research.

Journal of Verbal Learning and Verbal Behavior 11:671-684.

Craik, F. & E. Tulving (1975), Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104:268-294.

Cross, D. (1982), Aural discrimination and achievement. *The British Journal of Language Teaching* 20:139-143.

Daan, J. (1969), Bilingualism of Dutch immigrants in the U.S.A. In: A. Graur (ed.), *Actes du Xe Congrès International des Linguistes I.* Bucarest: A.R.S.R., 759-763.

Daan, J. (1971), Bilingualism of Dutch immigrants in the U.S.A. In: V. Lange & H.-G. Roloff (eds.), *Dichtung, Sprache, Gesellschaft.* Frankfurt: Athenäum (*Akten des IV. Internationalen Germanisten-Kongresses 1970 in Princeton*), 205-213.

Daan, J. (1987), *Ik was te bissie. Nederlanders en hun taal in de Verenigde Staten.* Zutphen: Walburg Pers.

Day, R. (1985), The use of the target language in context and second language proficiency. In: S. Gass & C. Madden (eds.), *Input in second language acquisistion.* Rowley, Mass.: Newbury House, 257-271.

Davies, A. (1986), Language loss and symbolic gain: The meaning of institutional maintenance. In: B. Weltens et al. (eds.), 117-127.

De Bot, K., P. Gommans & C. Rossing (1988), L1 loss in an L2-environment: Dutch immigrants in France. To appear in: H. Seliger & R. Vago (eds.).

De Bot, K. & T. Lintsen (1986), Foreign-language proficiency in the elderly. In: B. Weltens et al. (eds.), 131-141.

De Bot, K. & R. Schreuder (1987), Taalverlies en taalherstel: Lexicale vaardigheden in het Frans als vreemde taal. Research proposal submitted to the University of Nijmegen.

De Bot, K. & B. Weltens (1985), Taalverlies: Beschrijven versus verklaren. *Handelingen van het 38e Nederlands Filologencongres*, 51-61.

De Bot, K. & B. Weltens (1988), Recapitulation, regression, and language loss. To appear in: H. Seliger & R. Vago (eds.).

Dixon, W. (ed.) (1983), *BMDP Statistical software.* Berkeley, CA: University of California Press.

Dorian, N. (1977), The problem of the semi-speaker in language death. *Linguistics* 191:23-32.

Dorian, N. (1981), *Language death: The life cycle of a Scottish Gaelic dialect.* Philadelphia: University of Pennsylvania Press.

Dressler, W. & R. Wodak-Leodolter (1977), Language preservation and language death in Brittany. *Linguistics* 191:33-44.

Ebbinghaus, H. (1885), *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie.* Leipzig: Duncker & Humblot.

Edwards, G. (1976), Second-language retention in the Canadian public service. *The Canadian Modern Language Review* 32:305-308.

Edwards, G. (1977), *Second language retention in the public service of Canada.* Ottawa: Public Service Commission of Canada.

Efron, B. (1979), Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7:1-26.

Eggermont, J. & S. Hoekstra (1975), *Grammaire fondamentale*. Zutphen: Thieme.

Ervin-Tripp, S. (1974), Is second language learning like the first? *TESOL Quarterly* 8:111-127.

Evers, R. (1980), Zelfbeoordeling van taalvaardigheid Frans. *Toegepaste Taalwetenschap in Artikelen* 7:198-217.

Ferguson, G. (1981), *Statistical analysis in psychology and education*. New York: McGraw-Hill (5th ed.).

Flaugher, R. & R. Spencer (1967), College foreign language placement and the intervening years problem. *Modern Language Journal* 51:394-398.

Freed, B. (1980), The problem of language skill loss. Paper presented at the Annual Meeting of the Modern Language Association, New York, December 30, 1980.

Freed, B. (1982), Language loss: Current thoughts and future directions. In: R. Lambert & B. Freed (eds.), 1-5.

Fuller, E. (1982), A study of Navajo language maintenance and shift. Diss. University of Pittsburgh (*Dissertation Abstracts International* 43:706-A).

Galbraith, N. (1981), A study of transfer in language loss. Some problems and considerations. Master's thesis University of Utrecht.

Gardner, R. (1982), Social factors in language retention. In: R. Lambert & B. Freed (eds.), 24-43.

Gardner, R., R. Lalonde & J. MacPherson (1985), Social factors in second-language attrition. *Language Learning* 35:519-540.

Gardner, R., R. Lalonde, R. Moorcroft & F. Evers (1987), Second language attrition: The role of motivation and use. *Journal of Language and Social Psychology* 6:29-47.

Gardner, R. & W. Lambert (1959), Motivational variables in second language acquisition. *Canadian Journal of Psychology* 13:266-272.

Gardner, R. & W. Lambert (1972), *Attitudes and motivation in second language learning*. Rowley, Mass.: Newbury House.

Geoghegan, B. (1950), The retention of certain secondary school subjects by high school students over the summer vacation period. Diss. Fordham University.

Ginjaar-Maas, N. (1985), *Het voorbereidend hoger onderwijs*. The Hague: Staatsuitgeverij (Tweede Kamer, vergaderjaar 1985-1986, 19 114, nos. 2-3).

Ginsberg, R. (1986), Issues in the analysis of language loss: Methodology of the Language Skills Attrition Project. In: B. Weltens et al. (eds.), 19-36.

Godsall-Myers, J. (1981), The attrition of language skills in German classroom bilinguals - A case study. Diss. Bryn Mawr College.

Gonzo, S. & M. Saltarelli (1983), Pidginization and linguistic change in emigrant languages. In: R. Andersen (ed.), *Pidginization and creolization as language acquisition*. Rowley, Mass.: Newbury House, 181-197.

Gras, R. (1967), *Studietoetsen voor moderne talen*. Groningen: Wolters-Noordhoff.

Gregg, V. (1975), *Human memory*. London: Methuen.

Grendel, M., H. Kneepkens, C. Kuijpers & J. Poppe (1985), Taalvaardigheid in het Frans als vreemde taal: Grammaticale aspecten. Unpublished MS Department of Applied Linguistics, University of Nijmegen.

Grendel, M. & J. Poppe (1986), Grammaticale aspecten van het Frans als vreemde taal.

Master's thesis University of Nijmegen.

Groot, P. (1976), *Luistervaardigheid Frans-Duits-Engels. Doelstelling en toetsing.* Amsterdam: Meulenhoff Educatief.

Gruneberg, M., P. Morris & R. Sykes (eds.) (1988), *Practical aspects of memory: Current research and issues, Volume 1.* Chichester: Wiley & Sons.

Guntenaar, N. & S. Willemsen (1987), J'ai fait mon ***dernier*** examen. Ik heb mijn ..... examen gedaan. Recherche sur la perte de mots français. Master's thesis Free University of Amsterdam.

Gussenhoven, C. (1981), Measuring the acceptability of voiced fricatives in Dutch. *Proceedings of the Institute of Phonetics* 5:96-129.

Hagen, A. & H. Münstermann (1985), Functieverlies en structuurverlies bij minderheidstalen en dialecten. *Handelingen van het 38e Nederlands Filologencongres*, 63-84.

Hagiwara, P. (1983), Student placement in French: Results and implications. *Modern Language Journal* 67:23-32.

Hammer, P. & M. Monod (1976), *English-French cognate dictionary.* Edmonton, Alberta: University of Alberta.

Hansen, L. (1980), Learning and forgetting a second language: The acquisition, loss and re-acquisition of Hindi-Urdu negative structures by English-speaking children. Diss. University of California, Berkeley.

Haugen, E., J. McClure & D. Thomson (eds.) (1981), *Minority languages today.* Edinburgh: University Press.

Hellström, S.-G. & S. Johansson (1969), *On parle français.* Groningen: Wolters-Noordhoff.

Heurlin, K. (1972), *Vive le français.* Zutphen: Thieme.

Hill, J. & K. Hill (1977), Language death and relexification in Tlaxcalan Nahuatl. *Linguistics* 191:55-69.

Hinofotis, F. & B. Snow (1980), An alternative cloze testing procedure: multiple-choice format. In: J. Oller & K. Perkins (eds.), *Research in language testing.* Rowley, Mass.: Newbury House, 129-133.

Hoppenbrouwers, C. (1982), Language change. A study of phonemic and analogical change with particular reference to S.E. Dutch dialects. Diss. University of Groningen.

Howard, G. & P. Dailey (1979), Response-shift bias. A source of contamination of self-report measures. *Journal of Applied Psychology* 64:144-150.

Hyltenstam, K. & C. Stroud (1985), The psycholinguistics of language choice and code-switching in Alzheimer's dementia: Some hypotheses. In: Å. Viberg (ed.), *Bilingualism and second language acquisition.* Stockholm: University of Stockholm (*Scandinavian Working Papers on Bilingualism* 1985, 4), 26-44.

Jakobson, R. (1941), *Kindersprache, Aphasie und allgemeine Lautgesetze.* Uppsala: Almqvist & Wiksell. (Published in English in 1968 as: *Child language, aphasia and phonological universals.* The Hague: Mouton.)

Jamieson, P. (1980), The pattern of urban language loss. *Australian and New Zealand Journal of Sociology* 16:102-109.

Janssen-van Dieten, A. & C. Raymakers-Volaart (1986), Verslag evaluatie examens

Nederlands 1986. Unpublished MS Department of Applied Linguistics, University of Nijmegen.

Jaspaert, K. & S. Kroon (1987), The relationship between global language proficiency tests and language loss. In: F. Beukema & P. Coopmans (eds.), *Linguistics in the Netherlands 1987*. Dordrecht/Providence: Foris, 91-100.

Jaspaert, K., S. Kroon & R. van Hout (1986), Points of reference in first-language loss research. In: B. Weltens et al. (eds.), 37-49.

Jochems, W. & F. Montens (1987), De multiple-choice cloze-toets als algemene taalvaardigheidstoets. *Tijdschrift voor Onderwijsresearch* 12:133-143.

Jonz, J. (1988), Constraints on cloze response. Paper presented at the 22nd Annual Tesol Convention, Chicago, March 7-12, 1988.

Jordens, P., K. de Bot, C. van Os & J. Schumans (1986), Regression in German case marking. In: B. Weltens et al. (eds.), 159-176.

Juilland, A. (1965), *Dictionnaire inverse de la langue française*. The Hague: Mouton.

Kennedy, L. (1932), The retention of certain Latin syntactical principles by first and second year Latin students after various time intervals. *Journal of Educational Psychology* 23:132-146.

Kerkman, H. (1982), Taalpsychologisch onderzoek naar de organisatie van het tweetalige lexicon: Een overzicht. *Gramma* 6:199-214.

Kerkman, H. (1984), Woordherkenning in twee talen. In: A. Thomassen, L. Noordman & P. Eling (eds.), *Het leesproces*. Lisse: Swets & Zeitlinger, 139-152.

Kerkman, H. (in prep.), Second-language acquisition and the mental lexicon. Diss. University of Nijmegen.

Kleijn, P. (1977), Uitspraakcursus Frans. Unpublished MS Department of French, University of Nijmegen.

Klein-Braley, C. (1985), Reduced redundancy as an approach to language testing. In: C. Klein-Braley & U. Raatz (eds.), *Fremdsprachen und Hochschule*. Bochum: Ruhruniversität (*AKS-Rundbrief* 13/14), 1-13.

Knibbeler, W. (1977), *Frans van Nederlanders. Toetsing van de tussentaalhypothese*. The Hague: Staatsuitgeverij (*SVO reeks* 5).

Kolk, H. (1974), Experiments in forgetting. Diss. University of Nijmegen.

Koster, C. (1987), Word recognition in foreign and native language. Effects of context and assimilation. Diss. University of Utrecht.

Kuhlemeier, H. & R. van Werkhoven (1984), Inventarisatie van leerboeken talen in de onderbouw. *Levende Talen* 389:98-101.

Lambert, R. & B. Freed (eds.) (1982), *The loss of language skills*. Rowley, Mass.: Newbury House.

Lambert, R. & S. Moore (1984), Recent research on language skill attrition. *ERIC/CLL News Bulletin* 8:1-8.

Lambert, R. & S. Moore (1986), Problem areas in the study of language attrition. In: B. Weltens et al. (eds.), 177-184.

Lapkin, S. & M. Swain (1977), The use of English and French cloze tests in a bilingual education program evaluation: Validity and error analysis. *Language Learning* 27:279-314.

Lee, Y. (1985), Investigating the validity of the cloze score. In: Y. Lee, A. Fok, R. Lord & G. Low (eds.), *New directions in language testing.* Oxford: Pergamon, 137-147.

Levelt, W. (1978), Skill-theory and language teaching. *Studies in Second Language Acquisition* 1:53-70.

Lightbown, P. & N. Spada (1987), Learning English in intensive programs in Quebec schools: 1986-87. Unpublished MS, Concordia University/McGill University.

Loftus, G. & E. Loftus (1976), *Human memory: The processing of information.* Hillsdale, NJ: Erlbaum.

Luijten, A. (1984), *Eindexamens HAVO/VWO 1984.* Arnhem: Cito (*General Publication* 38).

Maas-de Brouwer, T. (1983), Normeren van centrale examens. In: W. Weeda (ed.), *Examens in discussie.* Groningen: Wolters-Noordhoff, 173-185.

Macht, K. & F. Steiner (1983), *Erfolgsfaktoren des Vokabellernens. Untersuchung zum aktiven Englischen Wortschatz von Hauptschulabgängern.* Augsburg: University of Augsburg (*Augsburger I- & I-Schriften* 25).

Markham, P. (1987), Rational deletion cloze processing strategies: ESL and native English. *System* 15:303-311.

Mattheier, K. (1986), Dialektverfall, Dialektabbau, Dialektveränderung. *Zeitschrift für Literaturwissenschaft und Linguistik* 62:58-73.

McMahon, N. (1946), The effects of summer vacation on retention of Latin vocabulary. Master's thesis Fordham University.

MEERLING (1981), *Methoden en technieken van psychologisch onderzoek. Deel 2: Data-analyse en psychometrie.* Meppel: Boom.

Messelink, J. & H. Verkuylen (1984), Het verlies van woordenkennis van het Frans. Master's thesis University of Nijmegen.

Moorcroft, R. & R. Gardner (1987), Linguistic factors in second language loss. *Language Learning* 37:327-340.

Münstermann, H. & T. Hagen (1986), Functional and structural aspects of dialect loss: A research plan and some first results. In: B. Weltens et al. (eds.), 75-96.

Naveh-Benjamin, M. (1988), Retention of cognitive structures learned in university courses. In: M. Gruneberg et al. (eds.), *Practical aspects of memory: Current research and issues, Volume 2.* Chichester, Wiley & Sons, 383-388.

Neisser, U. (1984), Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General* 113:32-35.

Nicholas, M., L. Obler & M. Albert (1982), Lexical access in healthy aging and in Alzheimer's dementia. Paper presented at the Boston University Language Development Conference, October 10, 1982.

Nienhuis, L. (1977), Het toetsen van spreekvaardigheid. Een onderzoek uitgevoerd voor het Frans ten behoeve van het schoolonderzoek moderne vreemde talen. The Hague: Staatsuitgeverij (*SVO reeks* 8).

Nyssen, R. & S. Crahay (1960), Etude des capacités de définition et d'évocation des mots en fonction de l'age. *Acta Psychologica* 17:1-23.

Obler, L. (1982), Neurolinguistic aspects of language loss as they pertain to second language attrition. In: R. Lambert & B. Freed (eds.), 60-79.

Oller, J (1973), Cloze tests of second language proficiency and what they measure. *Language Learning* 22:105-118

Olshtain, E (1986), The attrition of English as a second language: A case of Hebrew-speaking children In. B. Weltens et al (eds), 185-202.

Oskarsson, M (1984), *Self-assessment of foreign language skills: A survey of research and development work* Strasbourg: Council of Europe.

Oud-de Glas, M (1985), Het Nederlandse vreemde-talenonderwijs sinds de Mammoetwet In· A van Wieringen (ed), *Beeld van het voortgezet onderwijs.* Assen: Van Gorcum, 21-31.

Oxford, R. (1982a), Research on language loss - A review with implications for foreign-language teaching *Modern Language Journal* 66.160-169.

Oxford, R. (1982b), Technical issues in designing and conducting research on language skill attrition In. R Lambert & B Freed (eds.), 119-137

Pan, B & J Berko-Gleason (1986), The study of language loss. Models and hypotheses for an emerging discipline *Applied Psycholinguistics* 7 193-206

Pauwels, A (1986), Diglossia, immigrant dialects and language maintenance in Australia The case of Limburgs and Swabian. *Journal of Multilingual and Multicultural Development* 7·13-30.

Popham, W (1978), *Criterion-referenced measurement* Englewood Cliffs, NJ. Prentice-Hall.

Postman, L. (1971), Transfer, interference and forgetting. In. J. Kling & L. Riggs (eds.), *Woodworth & Schlosberg's experimental psychology.* New York: Holt, Rinehart & Winston (3rd ed.), 1019-1132.

Pratella, W (1969), The retention of first and second year Spanish over the period of the summer vacation. Diss. Fordham University (*Dissertation Abstracts International* 31, 1970-71.235-a).

Py, B. (1986), Native language attrition amongst migrant workers: Towards an extension of the concept of interlanguage. In: E. Kellerman & M. Sharwood Smith (eds), *Crosslinguistic influence in second language acquisition.* Oxford· Pergamon, 163-172.

Rea, C & V Modigliani (1988), Educational implications of the spacing effect. In· M. Gruneberg et al (eds.), 402-406

Robison, R. (1985), The effect of the summer vacation on language attrition in secondary school students of first-year Spanish. Diss Ohio State University (*Dissertation Abstracts International* 46, 1985:637-a)

Romaine, S (1986), Sprachmischung und Purismus: Sprich mir nicht von Mischmasch. *Zeitschrift für Literaturwissenschaft und Linguistik* 62:92-107

Savard, J -G & J Richards (1970), *Les indices d'utilité du vocabulaire fondamental français* Québec: Les Presses d'Université Laval.

Sawaie, M (1986), The present and future status of a minority language· The case of Arabic in the United States. *Journal of Multilingual and Multicultural Development* 7 31-40

Schaie, K (1965), A general model for the study of developmental problems *Psychological Bulletin* 64 92-107.

Scherer, G (1957), The forgetting rate in learning German *German Quarterly* 30 275-277

Schils, E (1988), Bootstraps, een nieuw statistisch alternatief? Paper presented at the ANeLA Studiedag, Utrecht, March 26, 1988 Forthcoming in *Toegepaste Taalwetenschap in Artikelen* 31

Schils, E & M van der Poel (forthc), The reliability ritual Forthcoming in *Gramma*

Schils, E & F Reelick (1985), Woordutiliteit, woordfrequentie en woorddispersie *Gramma* 9 89-103

Schlieben-Lange, B (1977), The language situation in Southern France *Linguistics* 191 101-108

Schouten-van Parreren, C (1983), Wisseling van de wacht in de vreemdetalen didactiek? De receptief-handelingspsychologische benadering *Levende Talen* 378 22 29

Schumans, J , C van Os & B Weltens (1985), Vocabulairekennis in de vreemde taal na beeindiging van het onderwijs *Toegepaste Taalwetenschap in Artikelen* 23 81-89

Seliger, H (1985), Primary language attrition in the context of other language loss and mixing Unpublished MS Queens College, City University of New York

Seliger, H & R Vago (eds) (1988), *First language attrition Structural and theoretical perspectives.* Cambridge Cambridge University Press [Forthcoming]

Sharwood Smith, M (1983a), On explaining language loss In S Felix & H Wode (eds), *Language development at the crossroads* Tubingen Gunter Narr, 48-59

Sharwood Smith, M (1983b), On first language loss in the second language acquirer Problems of transfer In S Gass & L Selinker (eds), *Language transfer in language learning* Rowley, Mass Newbury House, 222-231

Smythe, P , G Jutras, J Bramwell & R Gardner (1973), Second language retention over varying intervals *Modern Language Journal* 57 400-405

Snow, M , A Padilla & R Campbell (1984), *Factors influencing language retention of graduates of a Spanish immersion program* Los Alamitos, CA National Center for Bilingual Research

Sprangers, M & J Hoogstraten (1988), On delay and reassessment of retrospective preratings *Journal of Experimental Education* 56 148-153

Stankovski, M (1982), The family language development of Serbo-Croatian/Croatian speaking children in Sweden aged 4-18 years Lund JUBA-project

Stanley, J (1971), Reliability In R Thorndike (ed) *Educational Measurement* Washington American Council on Education (2nd ed), 356-442

Stevens, G (1982), Minority language loss in the United States Diss University of Wisconsin (*Dissertation Abstracts International* 43 274-A)

Tabouret-Keller, A. & F Luckel (1981), Maintien de l'alsacien et adoption du français Eléments de la situation linguistique en milieu rural en Alsace *Langages* 61 39-62

Thomassen, A & G Kempen (1976), Geheugen In J Michon, E Eijkman & L de Klerk (eds), *Handboek der Psychonomie* Deventer Van Loghum Slaterus, 354-387

Thorndyke, P (1977), Cognitive structures in comprehension and memory of narrative discourse *Cognitive Psychology* 9 77-110

Trudgill, P (1983), *On dialect Social and geographical perspectives* Oxford Blackwell

Tsitsipis, L (1981), Language change and language death in Albanian speech communities in Greece A sociolinguistic study Diss University of Wisconsin (*Dissertation Abstracts International* 42 5176-A)

Valdman, A (1982), Language attrition and the administration of secondary school and college foreign language instruction In R Lambert & B Freed (eds), 155-175

Van Agt, A & S Wessels (1984), Uitspraakproblemen van het Frans bij Nederlanders Een perceptieve en produktieve uitspraaktoets Unpublished MS Department of Applied Linguistics, University of Nijmegen

Van Els, T (1981), From foreign language needs to educational policy *AILA Bulletin* 2 48-61

Van Els, T (1985), Errors and foreign language loss Paper presented at the FIPLV symposium Errors and Foreign Language Learning, London, September 2-6, 1985 [Forthcoming in R Freudenstein (ed), *Errors and foreign language learning* Oxford Pergamon]

Van Els, T (1986), An overview of European research on language attrition In B Weltens et al (eds), 3-18

Van Els, T , T Bongaerts, G Extra, C van Os & A Janssen-van Dieten (1984), *Applied linguistics and the learning and teaching of foreign languages* London Arnold

Van Els, T & W de Jong (1985), *Moderne vreemde talen* The Hague Wetenschappelijke Raad voor het Regeringsbeleid (*Werkdocumenten Basisvorming in het Onderwijs* 9)

Van Els, T & H Radstake (1987), *Het onderwijs van de moderne vreemde talen 1975-1985* Enschede Foundation for Curriculum Development

Van Els, T & B Weltens (1983), Taalverlies in het Frans als vreemde taal Research proposal submitted to the Dutch Linguistics Foundation

Van Els, T & B Weltens (1987), Foreign language loss research from a European point of view Paper presented at the 8th World Congress of Applied Linguistics, Sydney, August 16-21,1987 [Forthcoming in *ITL Review of Applied Linguistics* 1988]

Van Vlerken, M (1980), Adverbial placement in English A study of first language loss Master's thesis University of Utrecht

Van Weeren, J (1982), Spreektoetsen moderne vreemde talen Verantwoording van een methode *Levende Talen* 377 870-887

Veldman, F (1975), Tweetaligheid en interferentie *Driemaandelijksche Bladen* 27 117-123

Verheesen, M (1984), Moedertaalverlies bij Duitse immigranten in Nederland Unpublished MS Department of Applied Linguistics, University of Nijmegen

Verhoeven, L & R Boeschoten (1986), First language acquisition in a second language submersion environment *Applied Psycholinguistics* 7 241-256

Verkaik, P & P van der Wijst (1986), Taalverlies en woordherkenning in het Frans als vreemde taal Master's thesis University of Nijmegen

Warnant, L (1964), *Dictionnaire de la prononciation française* Gembloux Editions J

Duculot (2nd ed.).
Warnant, L (1973), *Dictionnaire des rimes orales et écrites.* Paris: Larousse.
Weis, D. (1986), Untersuchungen zur langfristigen Verfügbarkeit von Wortschatz im Leistungsfach Englisch. *Neusprachliche Mitteilungen aus Wissenschaft und Praxis* 39:174-180.
Weltens, B. (1987), The attrition of foreign-language skills: A literature review. *Applied Linguistics* 8:22-37.
Weltens, B., K. de Bot & T. van Els (eds.) (1986a), *Language attrition in progress.* Dordrecht/Providence: Foris.
Weltens, B., M. Grendel & J. Poppe (1986b), Beheersing van de Franse grammatica na het onderwijs. *Levende Talen* 415:575-579.
Weltens, B. & T. van Els (1986) The attrition of French as a foreign language: Interim results. In: B. Weltens et al. (eds.), 205-221.
Wijnstra, J. (1977), Het gebruik van de cloze procedure als maat voor schriftelijke taalbeheersing. *Tijdschrift voor Onderwijsresearch* 2:262-269.
Yorozuya, R. & J. Oller (1980), Oral proficiency scales: Construct validity and the halo effect. *Language Learning* 30:135-153.

LIST OF APPENDICES

Appendix 1: Multiple-choice cloze test

Instructions

First read the entire text carefully. Then try to find the right word for each gap. Tick this word on the answering sheet. In any case, tick an answer. Good luck!

Le métro

"J'habite à la Porte de Vincennes, et je travaille dans un bureau, rue La Fayette. Tous les matins, je ...1... le métro à huit ...2... vingt, à la station '...3... de Vincennes'.
Il y ...4... du monde, mais je ...5... fais pas la queue, ...6... j'ai toujours des ...7... dans mon sac. A ...8... station 'Châtelet', je change ...9... aller jusqu'à 'Chaussée d'..10... . J'arrive toujours à ..11... heure à mon bureau: ..12... commence mon travail à ..13... heures. Le soir, je ..14... du bureau avec mon ..15..., et nous prenons le ..16... à 'Chaussée d'Antin'. ..17... y a toujours une ..18... longue queue, et nous ..19... souvent quatre ou cinq ..20... avant de passer sur ..21... quai. Tous les bureaux ..22... à six heures et, ..23... six heures cinq, tous ..24... employés rentrent chez eux.

..25... n'aime pas prendre ..26... métro à cette heure-là. ..27... est toujours debout et, ..28... on veut descendre, on ..29... demander à trois ou ..30... personnes: "Est-ce que vous descendez, ..31..., Madame ou Mademoiselle?" avant ..32... arriver près de la ..33... . Quand on sort de ..34... station, on respire mieux!

..35... en métro n'est ..36... très agréable, mais c'..37... rapide et bon marché. ..38... un ticket de seconde ..39..., vous pouvez voyager dans ..40... toute une journée!".

Answering sheet "Le métro"

Tick the correct answer.

1. 0 prends
   0 prendre
   0 pris

2. 0 et
   0 moins
   0 heures

3. 0 Train
   0 Porte
   0 Paris

4. 0 as
   0 a
   0 est

5. 0 ne
   0 là
   0 jamais

6. 0 parce que
   0 quand
   0 mais

7. 0 tickets
   0 cartes
   0 francs

8. 0 cette
   0 la
   0 une

9. 0 d'
   0 pour
   0 métro

10. 0 Orly
    0 Avenue
    0 Antin

11. 0 l'
    0 une
    0 un

12. 0 et
    0 on
    0 je

13. 0 neuf
    0 deux
    0 trois

14. 0 fais
    0 sors
    0 quitte

15. 0 sac
    0 amie
    0 papiers

16. 0 trajet
    0 voyage
    0 métro

17. 0 Là
    0 Il
    0 Ils

18. 0 vrai
    0 beaucoup
    0 très

19. 0 entendons
    0 attendons
    0 sommes

20. 0 minutes
    0 métros
    0 heures

21. 0 le
    0 la
    0 les

22. 0 fermes
    0 fermer
    0 ferment

23. 0 en
    0 à
    0 de

24. 0 les
    0 nos
    0 l'

25. 0 Il
    0 Je
    0 Ils

26. 0 le
    0 du
    0 la

27. 0 On
    0 C'
    0 Il

28. 0 quand
    0 car
    0 pourquoi

29. 0 va
    0 doit
    0 veut

30. 0 quatre
    0 plus
    0 des

31. 0 Monsieur
    0 pardon
    0 encore

32. 0 l'
    0 d'
    0 qu'

33. 0 porte
    0 personne
    0 métro

34. 0 ce
    0 la
    0 métro

35. 0 Voyager
    0 Mais
    0 Et

36. 0 toujours
    0 parfois
    0 pas

37. 0 est
    0 aller
    0 était

38. 0 Avec
    0 Acheter
    0 Payer

39. 0 fois
    0 prix
    0 classe

40. 0 métro
    0 Paris
    0 lui

Instructions

First read the entire text carefully. Then try to find the right word for each gap. Tick this word on the answering sheet. In any case, tick an answer. Good luck!

Comment mangent les Français?

Comment mangent les Français? Leur petit déjeuner est vraiment petit! Le ...1..., les Français prennent seulement ...2... café au lait dans ...3... grand bol et des ...4... ou des croissants. Ah! ...5... croissants chauds du matin! ...6... les sent dans la ...7... quand on approche de ...8... boulangerie. Ils sentent bon ...9... ils sont délicieux! A ..10..., beaucoup de Parisiens n'..11... pas le temps de ..12... chez eux et préfèrent ..13... dans les restaurants et ..14... cafés près de leur ..15... . Pour beaucoup d'entre ..16..., le menu du déjeuner ..17... est souvent un sandwich ..18... une tasse de café.
Mais dans les petites ..19... et à la campagne, ..20... déjeuner est le plus ..21... repas de la journée: ..22... mange des hors-d'oeuvre, de ..23... viande, des légumes, du ..24..., des fruits ou un ..25...; on boit du vin, ..26... la bière ou de ..27... eau, mais jamais du ..28... avec les repas. Le ..29..., à sept heures ou ..30... huit heures, le dîner ..31... encore un gros repas, ..32... il y a souvent ..33... la soupe à la ..34... des hors-d'oeuvre.
On passe ..35... beaucoup de temps à ..36... en France: un bon ..37... est souvent le plus ..38... plaisir du dimanche. Demandez ..39... adresse d'un bon ..40... restaurant à vos amis ..41... . Ils en connaissent sûrement ..42... .

Answering sheet "Comment mangent les Français?"

Tick the correct answer.

1. 0 jour
   0 déjeuner
   0 matin

2. 0 une
   0 du
   0 avec

3. 0 un
   0 la
   0 pain

4. 0 cadets
   0 pains
   0 tartines

5. 0 les
   0 bons
   0 des

6. 0 Ils
   0 On
   0 Je

7. 0 rue
   0 cuisine
   0 main

8. 0 cette
   0 meilleure
   0 la

9. 0 mais
   0 et
   0 comme

10. 0 matin
    0 déjeuner
    0 midi

11. 0 ait
    0 ont
    0 a

12. 0 voyager
    0 rentrer
    0 rester

13. 0 mange
    0 mangent
    0 manger

14. 0 boivent
    0 prennent
    0 les

15. 0 bureau
    0 maison
    0 amis

16. 0 Ils
    0 eux
    0 leur

17. 0 grand
    0 c'
    0 midi

18. 0 à
    0 et
    0 chez

19. 0 villes
    0 restaurants
    0 villages

20. 0 un
    0 le
    0 son

21. 0 bien
    0 gros
    0 grande

22. 0 il
    0 je
    0 on

23. 0 la
    0 leur
    0 grand

24. 0 fromage
    0 pomme de terre
    0 frites

25. 0 pomme
    0 dessert
    0 tarte

26. 0 de
    0 jamais
    0 ou

27. 0 l'
    0 minérale
    0 fraîche

28. 0 limonade
    0 buvable
    0 lait

29. 0 dîner
    0 repas
    0 soir

30. 0 environ
    0 à
    0 pendant

31. 0 est
    0 était
    0 prend

32. 0 mais
    0 qu'
    0 parfois

33. 0 de
    0 après
    0 dans

34. 0 place
    0 repas
    0 carte

35. 0 très
    0 avec
    0 encore

36. 0 mangent
    0 café
    0 table

37. 0 truc
    0 repas
    0 chose

38. 0 joli
    0 beau
    0 grand

39. 0 un
    0 votre
    0 l'

40. 0 connu
    0 français
    0 petit

41. 0 français
    0 toujours
    0 vite

42. 0 les
    0 plusieurs
    0 bien

Appendix 2: Samples from the LC test.

(B,1) [After having given a short introduction on the subject of the panel discussion, and after having introduced the panel members, the interviewer asks his first question:]
Alors monsieur Quimelle nous a apporté un sondage de Télérama qui a été euh, fait auprès des adolescents de 12 à 17 ans, la question "Quel est euh, le nom de votre héros préféré?"

Alors parmi les noms cités euh, on relève cinq euh, personnages euh, du cinéma, cinq acteurs de cinéma français et américains et un personnage historique, qui est euh, Napoléon. Alors je crois que c'est assez frappant de constater cette euh, euh, (.....) d'acteurs de cinéma.

[The answering sheet reads:] 1. A absence totale
B grande proportion

(C,3) Christiaan Huygens est né à La Haye, le 14 avril 1629, il est né dans un milieu familial priviligié. Permettez-moi de vous évoquer brièvement l'enfance et la formation de Christiaan Huygens. D'abord n'oublions pas que Christiaan Huygens était le fils de Constantin, le grand poète néerlandais, qui a vu ses mérites personnels reconnus avec un poste important mais surtout influent: celui de secrétaire du Prince d'Orange. C'est dans ce milieu que Christiaan a certainement goûté les avantages de tous les contacts nationaux, mais aussi largement internationaux, que son père, Constantin, avait à la Cour de La Haye. En même temps, Christiaan Huygens a pu profiter de la formation d'un père, qui avait reçu, lui-même, une formation humaniste. Alors Christiaan Huygens a certainement eu sa première formation à la maison de son père, par l'entremise du père, mais aussi par de bons gouverneurs, de bons précepteurs invités chez les Huygens.

[The answering sheet reads:]
3. Que dit Mme Savatier au sujet de l'enfance de Huygens?
A Huygens doit beaucoup aux qualités de son père.
B Huygens a eu une jeunesse irrégulière à cause des voyages de son père.
C Huygens est longtemps resté dans l'ombre d'un père beaucoup plus célèbre.

Appendix 3: Samples from the RC test

V UNE ÉCOLE DE TOLÉRANCE

De nombreux psychosociologues américains ont observé les étudiants étrangers séjournant dans leur pays au cours des vingt dernières années. Largement convergentes, leurs conclusions ont sans doute une portée générale. Selon eux, la bonne adaptation d'un jeune étranger dépend d'abord de sa volonté d'assimilation et de son absence de préjugés. Il faut ensuite qu'il ait peu de contacts avec des compatriotes, et beaucoup d'occasions de relations amicales avec la population locale. Il doit enfin réussir de façon correcte sur le plan scolaire et avoir le sentiment que son pays est apprécié. Ainsi, les étudiants asiatiques et africains séjournant en Europe ou aux Etats-Unis éprouvent, parce que leurs pays sont mal connus et peu estimés, un sentiment d'infériorité culturelle qui leur fait adopter une attitude *a priori* défensive: leur susceptibilité "nationale" entraîne une forte susceptibilité personnelle.

Vivre dans un pays étranger, constatent les chercheurs américains, n'amène pas forcément à l'apprécier davantage. Une bonne adaptation modifie favorablement l'image qu'on avait du pays avant d'y séjourner, mais certains étudiants *a priori* hostiles à leur pays d'accueil peuvent repartir renforcés dans leurs convictions: soit que leur attitude leur ait attiré des expériences négatives, soit qu'ils ne retiennent que celles susceptibles de confirmer leurs convictions.

Un séjour à l'étranger de longue durée modifie la façon dont les adolescents voient le monde et les hommes. Il rend plus "libéral", bien que les convictions philosophiques ou religieuses soient peu affectées. Demeuré différent des individus du pays d'accueil, l'étudiant étranger le devient aussi de ses compatriotes. Ce sont les ressortissants des pays du tiers-monde ayant séjourné aux Etats-Unis qui ont le plus de mal à se réadapter.

Le docteur Cigdem Kagitçiba-Si, de l'université Bogazici d'Istanbul, a interrogé deux groupes de lycéens turcs sur leur attitudes et croyances avant leur départ pour les Etats-Unis en 1970 et 1971, puis à leur retour en 1971 et 1972, et comparé leurs réponses à celles de lycéens restés en Turquie. Les conclusions confirment les travaux américains. Il constate que les lycéens qui ont vécu aux Etats-Unis sont plus ouverts à la dimension internationale et plus tolérants, au minimum moins racistes, moins nationalistes et moins religieux. Et moins soumis à l'autorité de leurs parents ou des enseignants qu'avant leur départ. Mais ils ne sont pas moins attachés à leur famille et à leur pays.

Le Monde de l'Education, 1979

42

Les mots "Largement convergentes," (ligne 2) pourraient être remplacés par
A Dès qu'elles sont largement convergentes,
B Même si elles sont largement convergentes,
C Quoiqu'elles soient largement convergentes,
D Vu qu'elles sont largement convergentes,

50

Qu'est-ce que l'auteur veut dire par le titre?
A C'est à l'école qu'on devrait apprendre la tolérance.
B Le séjour à l'étranger d'un étudiant le rend plus tolérant.
C Les étudiants doivent s'assimiler mieux au pays d'accueil.
D Les pays d'accueil doivent être plus tolérants envers les étudiants étrangers.

Appendix 4: Oppositions tested in the PHO-LI test

	Initially	*Medially*	*Finally*
/p/-/b/:	pas-bas	happer-abbé	trompe-trombe
/t/-/d/:	temps-dans	roter-rôder	tarte-tarde
/k/-/g/:	quai-gai	vaquer-vaguer	roc-rogue
/f/-/v/:	faire-vert	café-caver	actif-active
/s/-/z/:	sel-zèle	ils sont-ils ont	dix-dise
/ʃ/-/ʒ/:	-- --	-- --	biche-bige

	Monosyllabic	*Disyllabic-1*	*Disyllabic-2*
/ɑ/-/ɔ/:	brasse-brosse	damner-donner	débâte-débotte
/ɑ/-/ɛ/:	pas-paix	baser-baiser	pourras-pourrais
/ɔ/-/ɛ/:	poste-peste	horreur-erreur	divorce-diverse
/ɑ̃/-/ɔ̃/:	temps-ton	penser-poncer	néant-néon
/ɑ̃/-/ɛ̃/:	banc-bain	tenter-teinter	enfant-enfin
/ɔ̃/-/ɛ̃/:	pont-pain	pompant-pimpant	retondre-reteindre

Fillers: aller; vous; bon; fois; ciment; blanche.

Appendix 5: Oppositions tested in the PHO-RE test

	Initially	*Medially*	*Finally*
/p/-/b/:	peau-beau	appris-abris	cap-cab
/t/-/d/:	toute-doute	coûter-couder	vite-vide
/k/-/g/:	coût-goût	écoute-égoutte	brique-brigue
/f/-/v/:	fous-vous	lofer-lover	neuf-neuve
/s/-/z/:	sain-zain	coussin-cousin	douce-douze
/ʃ/-/ʒ/:	-- --	-- --	marche-marge

	Monosyllabic	*Disyllabic-1*	*Disyllabic-2*
/ɑ/-/ɔ/:	basse-bosse	casser-cosser	empâte-empote
/ɑ/-/ɛ/:	mât-met	lasser-laisser	repas-repais
/ɔ/-/ɛ/:	sol-sel	former-fermer	adopte-adepte
/ɑ̃/-/ɔ̃/:	en-on	tendu-tondu	répand-répond
/ɑ̃/-/ɛ̃/:	pan-pain	enfant-infant	Provence-province
/ɔ̃/-/ɛ̃/:	ponte-pinte	poncer-pincer	galion-Galien

Fillers:	pot-peau	abri-abris	cap-cape
	tout-toux	couder-coudée	vide-vides
	coup-coût	écoute-écoutent	brick-brique
	vous-voue	lover-lové	neuf-neufs
	saine-seine	cousin-cousins	doue-doux
	-- --	-- --	marche-marchent
	bas-bât	cassé-casser	dot-dote
	mais-mes	aîné-henné	paît-paix
	celle-sel	hors-or	cette-sept
	en-an	temps-tant	pand-pend
	pain-pin	pante-pente	étain-éteint
	peinte-pinte	pincer-pincée	teinte-tinte

Appendix 6: Target words in the LEX-LI test

High-frequency cognates

(French)	(Dutch)	(English)
accord	akkoord	agree(ment)
forme	vorm	form
groupe	groep	group
moteur	motor	engine
papier	papier	paper
personne	persoon	person
plaisir	plezier	favour
simple	simpel	simple
table	tafel	table
triste	triest	sad

Low-frequency cognates

(French)	(Dutch)	(English)
action	actie	action
bébé	baby	baby
cercle	cirkel	circle
citron	citroen	lemon
masse	massa	mass
olive	olijf	olive
passion	passie	passion
pilote	piloot	pilot
signal	signaal	signal
tomate	tomaat	tomato

High-frequency non-cognates

(French)	(Dutch)	(English)
avion	vliegtuig	plane
bras	arm	arm
chien	hond	dog
enfant	kind	child
gare	station	(train) station
main	hand	hand
petit	klein	small
pied	voet	foot
soleil	zon	sun
terre	grond	ground

Low-frequency non-cognates

(French)	(Dutch)	(English)
barrage	wegversperring; dam	road block; dam
cheville	enkel	ankle
douceur	zachtheid	softness
espoir	hoop	hope
outil	gereedschap	utensil; tool
profond	diep	deep
reproche	verwijt	reproach
tiers	eenderde	one-third
tombeau	graf	grave
virgule	komma	comma

Appendix 7: Target words in the LEX-RE test

High-frequency cognates

(French)	(Dutch)	(English)
classe	klas	class
famille	gezin	family
juste	juist, goed	just, right
minute	minuut	minute
musique	muziek	music
normal	normaal	normal
porte	poort; deur	gate, door
province	provincie	province
public	publiek	audience
vacances	vakantie	holiday

Low-frequency cognates

(French)	(Dutch)	(English)
avocat	advocaat	lawyer
canal	kanaal	chanel
finances	financien	finance
fontaine	fontein	fountain
logique	logisch	logical
palais	paleis	palace
planche	plank	board, shelf
portrait	portret	portrait
prince	prins	prince
salaire	salaris	salary

High-frequency non-cognates

(French)	(Dutch)	(English)
argent	geld	money
bateau	boot	boat
chaud	warm	hot
cuisine	keuken	kitchen
dernier	laatste	last
jardin	tuin	garden
malade	ziek	ill
pain	brood	bread
repas	maaltijd	meal
robe	jurk	dress

Low-frequency non-cognates

(French)	(Dutch)	(English)
boisson	drank	drink
caprice	gril	whim
drap	laken	sheet
évier	gootsteen	sink
grève	staking	strike
mouche	vlieg	fly
mouton	schaap	sheep
puissant	machtig	powerful
sable	zand	sand
vigne	wijngaard	vineyard

Appendix 8: Phenomena tested in the MS-LI/MS-RE tests

		ITEM TYPE*
1. Pers. pronoun: dir./indir. obj., sing.,	*le/lui/la***	C2
2. Pers. pronoun: dir./indir. obj., plur.,	*les/leur/leurs*	S
3. Pers. pronoun: after prep., fem., sing.,	*lui/elle/la*	S
4. Pers. pronoun: the contrast *en/y*,	*en/y/là*	S
5. Rel. pronoun: after prep., masc., sing.,	*quel/lequel/qui*	S
6. Rel. pronoun: subject/object,	*qui/que/quel*	C2
7. Interrog. pronoun: subject/object,	*qui/que/est-ce que*	S
8. Indef. pronoun: sing./plur. ("all the ..."),	*toute/tous/toutes*	S
9. Indef. pronoun: sing./plur. ("every"),	*tout/tous/toute*	S
10. Indef. pronoun: subject/object,	*ce qui/ce que/que*	C2
11. Indef. pronoun: negative, personal,	*personne/rien/aucun*	S
12. Indef. pronoun: negative, non-personal,	*rien/personne/aucun*	S
13. Poss. pronoun: 'possession-agreement',	*son/sa/la*	C3
14. Dem. pronoun: masc., sing.,	*ce/cette/cet*	C3
15. Dem. pronoun: fem., sing.,	*cette/ce/cet*	C3
16. Dem. pronoun: substantive use,	*celui/ce/celle*	C3
17. Refl. pronoun: 1st person sing.,	*nous/se/Ø*	C3
18. Refl. pronoun: 3rd person plur.,	*se/leur/Ø*	S
19. Def. article: plur., generic reference,	*les/des/Ø*	C3
20. Part. article: fem., sing.,	*de la/de/Ø*	C1
21. Part. article: after quantifier,	*de/des/Ø*	C1
22. Preposition: after verb,	*Ø/de/à*	S
23. Preposition: "in" (place),	*dans/en/entre*	S
24. Preposition: "with"	*chez/à/en*	S
25. Adjective: gender/number marking,	*neuves/neufs/neuf*	C2
26. Adverb: formation from adjective,	*lentement/lente/lent*	C2
27. Verb: infinitive after auxiliary,	*passer/passez/passé*	S
28. Verb: choice of *avoir/être*,	*est/a/peut*	S
29. Verb: 1st/3rd p. sing., simple present,	*écrit/écris/écrive*	S
30. Verb: plural, simple present, *être*,	*sommes/êtes/sont*	C2
31. Verb: plural, simple present,	*-ons/-ez/-ent*	C2
32. Verb: 3rd p. sing., simple past,	*avait/a/ait*	S
33. Verb: conditional past,	*aurait/aura/a*	S
34. Verb: imperative sing.,	*lève/levons/levez*	S
35. Verb: indicative/subjunctive,	*est/soit/ait*	S
36. Verb: subj./indic., 1st p. plur.,	*soyons/sommes/serons*	C1
37. Verb: subj./indic., 3rd p. sing.,	*vienne/vient/viendra*	C1
38. Verb: *avoir* as lex. verb with age,	*a/est/devient*	C3

39.	Verb: exist. sent., plur., simple past,	*il y avait/avaient/étaient*	C1
40.	Verb: past part., gender/number marking,	*prise/pris/prises*	C2

* S = similarity item; C = contrasting item: C1 = 'absent'; C2 = 'form/function'; C3 = 'distribution' (cf. section 3.7).

** The first alternative is the correct one. In cases where there is a dichotomous opposition, the second alternative is the second 'member', and the third is from a different word class, tense, etc.

Appendix 9: Can-do scales for listening and reading comprehension

We would like you to indicate how much difficulty it would cost you to carry out certain activities in French. We would like you to indicate this for two points in time, viz. (A) *at the moment*, and (B) *at the end of your training in French at school*. For each of the activities listed below, tick in column A how much difficulty they would present *at the moment*, and in column B how much much difficulty they would have presented *at the end of your training in French at school*. Please use the following scale:

1 = not at all
2 = with extreme difficulty
3 = with a lot of difficulty
4 = with some difficulty
5 = with little or no difficulty

LISTENING COMPREHENSION

		Mean (s.d.)
a.	In a personal conversation with a Frenchman, understand simple sentences like "Hello", "What's your name?", "Where do you live?".	4.55 (0.66)
b.	In a personal conversation, understand a Frenchman who speaks slowly and carefully, i.e. who deliberately adapts his/her speech to suit me.	3.81 (0.82)
c.	On the telephone, understand a Frenchman who speaks slowly and carefully, i.e. who deliberately adapts his/her speech to suit me.	3.25 (0.94)
d.	In face-to-face conversation with a Frenchman who speaks slowly and carefully, tell whether (s)he is referring past, present or future events.	3.39 (1.04)
e.	In face-to-face conversation, understand a Frenchman who is speaking as quickly and as colloquially to me as (s)he would do to another Frenchman.	2.06 (0.88)
f.	Understand films without subtitles.	2.84 (0.95)
g.	Understand news reports on the radio.	2.19 (0.98)
h.	On the radio, understand the words of a song I haven't heard before.	2.33 (1.06)

		Mean (s.d.)
i.	Understand sports reports (e.g. a soccer match) on the radio.	2.10 (1.10)
j.	Understand two Frenchmen when they are talking rapidly with each other.	1.51 (0.69)
k.	On the telephone, understand a Frenchman who is speaking as rapidly and as collquially as (s)he would to another Frenchman.	1.51 (0.73)

READING COMPREHENSION

a.	Read personal letters written to me in which the writer has deliberately used simple words and constructions.	4.21 (0.90)
b.	Read, on store fronts, the type of store (e.g. "dry cleaning", "book store", "butcher").	4.28 (0.83)
c.	Understand newspaper headlines.	3.87 (0.80)
d.	Read personal letters written as they would be to a Frenchman.	2.88 (0.98)
e.	Read magazine articles at a level such as those found in "Time" or "Newsweek" [in fact: "Elseviers Weekblad" and "De Haagse Post"] without using a dictionary.	2.07 (0.98)
f.	Read popular novels without using a dictionary.	2.53 (1.07)
g.	Read newspaper "want ads", even when many abbreviations are used.	2.12 (1.00)
h.	Read highly technical articles in a particular field, making little or no use of a dictionary.	1.72 (0.88)

Appendix 10: Comparison of ANOVA and ANCOVA results for the cloze test

The comparison reported here is the one between a straightforward ANOVA and an ANCOVA using the average school mark for French as a covariate. How small the difference between the two is, already becomes apparent from an inspection of the cell means resulting from either analysis:

	ANOVA		ANCOVA	
	Unadj. means	Unadj. loss	Adj. means	Adj. loss
Educ-6, Non-use-0	69.64		68.84	
		-1.68		-1.34
Educ-6, Non-use-4	71.32		70.18	
Educ-4, Non-use-0	62.32		63.51	
		1.32		1.76
Educ-4, Non-use-4	61.00		61.75	

This becomes even clearer from the complete table of results; here is the straightforward ANOVA:

SOURCE	SS	DF	MS	F
Education	1945	1	1945	55.12
Non-use	1	1	1	0.02
E x N	57	1	57	1.61
Error	3387	96	35	

The equivalent results of the ANCOVA were as follows:

SOURCE	SS	DF	MS	F
Education	937	1	937	29.39
Non-use	1	1	1	0.03
E x N	61	1	61	1.90
School mark	359	1	359	11.27
Error	3027	95	32	

The only difference between the two analyses is the effect of Education, which is reduced by the introduction of the covariate. This is not surprising, of course, since students partly base their choice of exam subjects on their school marks (cf. section 3.11); in other words, students choosing French tend to be somewhat better learners of French than those who drop it. Controlling the results for French school marks, therefore, implies levelling out to a certain degree the difference between the two training levels. Note, however, that even after controlling for school marks, the effect of Education is enormous (the critical value of F at the 1% level is only 6.96 in this case).

Now, one might doubt the adequacy of our choice of covariate, since its introduction into the analysis does not really affect the results. Interestingly enough, though, the analysis at the same time tells us we *have* picked a relevant covariate of test performance, judging by its F-ratio. Apparently, the groups do not differ very much in terms of school performance in French.

Appendix 11: Correlations between tests

	Cloze	LC	RC	PHO-LI	PHO-RE	LEX-LI	LEX-RE	MS-LI	MS-RE
Cloze	1	.70	.71	.25	.37	.69	.61	.69	.66
LC		1	.83	.30	.33	.65	.61	.69	.63
RC			1	.26	.34	.69	.60	.66	.60
PHO-LI				1	.48	.41	.05	.27	.12
PHO-RE					1	.45	.15	.32	.21
LEX-LI						1	.61	.56	.40
LEX-RE							1	.55	.52
MS-LI								1	.82
MS-RE									1

SUMMARIES

Summary (English)

The research project reported on here was concerned with the retention of school-learned French across a period of two and four years after course completion, in which there was hardly any exposure to the target language - hence the designation 'period of non-use'. The subjects were students who had had four or six years of training in French at the highest level of general secondary education in the Netherlands, *VWO*. The average number of hours of French instruction in this type of school is three per week; in other words, the students had had about 400 and 600 hours of French training respectively.

The combination of two training levels and three points in time resulted in a design with six groups, which was investigated in a combination of longitudinal and cross-sectional measurements (the arrows indicate longitudinal measurements):

		Years of non-use				
		0		2		4
Years of training	6	A		B	-->	C
	4	D	-->	E		F

Each group consisted of 25 subjects, who were given a number of *receptive* language tests and a questionnaire which contained a number of self-assessment scales. The tests were, on the one hand, global tests, i.e. tests of general (receptive) proficiency, and listening and reading proficiency; on the other hand, tests of phonological, lexical, and grammatical competence were administered.

The phonological, lexical, and grammar tests all contained specific sub-categories of items, in order to investigate whether specific sub-classes of elements or rules would be more susceptible to attrition than others. On all three linguistic levels the relation between native and target language - i.e. the absence or presence of contrast between elements or rules in the two

languages - was incorporated. The items in the phonological tests were further subdivided into vocalic and consonantal; the items in the lexical tests into high and low frequency.

The self-assessments consisted of global self-reports of pronunciation, vocabulary and grammar, and *can-do scales* for listening and reading comprehension (11 and 8 items respectively). Where applicable - i.e. in the case of groups B, C, E, and F - subjects were not only asked to give self-assessments of their *present* proficiency, but also of their proficiency *at the time when they completed their French courses*, so-called *retrospective* self-assessments. All scales used were five-point scales with all points marked.

The groups that were compared cross-sectionally were matched as well as possible on potentially relevant background variables, such as learning career, school marks for French, amount of Latin instruction, and out-of-school contact with French.

The results of all the self-assessment measures indicated heavy attrition, according to a remarkably consistent pattern: the amount of attrition reported was of about the same size for each of the two training levels; the attrition occurred in the first two years of non-use, and there was no further attrition in the second interval of two years. The global self-assessments suggested that - in the subjects' own perception - pronunciation suffered least, and grammar suffered most.

The test results, on the other hand, revealed quite a different picture. General (receptive) proficiency remained virtually unchanged during four years following training, and listening and reading proficiency even increased significantly over time. The three linguistic levels produced divergent results: phonology appeared to improve, vocabulary remained relatively stable, and grammar decreased.

Nevertheless, the influence of contrast between native language and target language could be attested on all three levels. For phonology, the category 'contrasting consonantal' (as in the minimal pairs *écoute-égoutte* and *neuf-neuve*) appeared to be the most difficult one, but at the same time it was the category that contributed most substantially to the overall improvement across the years. On the other hand, although the overall scores increased significantly, we were able to identify one particular set of items that exhibited some attrition, at least in the written phonology test. Again, contrast was involved, for it was the 'contrasting vocalic' opposition

between the nasal vowel in, for example, *répand* as opposed to the one in *répond*, that became increasingly problematic.

In the lexical tests, although they did not reveal any serious attrition, contrast played an important role as well, in this case in interaction with frequency. The non-cognates of low frequency, i.e. words such as *évier*, *mouche*, *vigne*, *caprice*, and *grève*, were by far the most difficult category of words, and they were also the words that exhibited some attrition.

The grammar tests, then, revealed losses of about 10-15% of the original knowledge, but the attrition again concentrated in the contrastive category. Moreover, we were able to identify the contrastive aspects of the French pronominal system as an area of grammar particularly susceptible to attrition.

The fact that very little attrition was found may be explained by the fact that we were dealing with subjects of relatively high levels of competence in French, certainly when compared to the subjects in many American investigations. In terms of learning theory, one could argue that all subjects had studied French in a context that psychologists would characterize as 'meaningful learning'. And it is a well-established fact that meaningful learning is much less subject to forgetting than rote learning. Moreover, if there is such a thing as a "critical threshold during learning", or a "critical mass of language", it may well be that our subjects had indeed reached such a level of mastery of French, making their French language skills relatively immune against forgetting.

None the less, there are strong indications that lexical and certainly grammatical skills are subject to attrition in the time interval investigated. In the case of lexical skills, further evidence was derived from a pilot experiment in which retrieval *speed*, rather than retrieval *success* was measured. The latter evidence suggests that the fact that our tests were administered without any time pressure - the aural tests were all leisurely paced; the written versions were even self-paced - may have obscured the real seriousness of the attrition that our subjects have experienced. Such a position would also explain why the subjects themselves could, at the same time, be so much more pessimistic about the retention of their French language skills.

Résumé (French)

Le projet de recherche, dont voici le compte-rendu, avait comme objet le maintien du français (appris à l'école) durant une période de deux et quatre années après l'achèvement de l'apprentissage du français, période pendant laquelle la langue cible était peu utilisée - d'où la désignation "période de non-emploi". Les sujets étaient des élèves ayant eu quatre ou six années de français au plus haut niveau dans l'enseignement secondaire aux Pays-Bas, le VWO. Le nombre moyen d'heures de français pour ce type d'enseignement est de trois heures par semaine; en d'autres termes, ils avaient eu respectivement 400 et 600 heures d'instruction du français.

La combinaison de deux niveaux d'instruction et de trois périodes différentes avait pour résultat une structure comportant six groupes. Ces groupes étaient étudiés à l'aide d'une combinaison de mesurages longitudinaux et 'cross-sectionelles' (les flèches indiquent les mesurages longitudinaux):

		Années de non-emploi				
		0		2		4
Années de	6	A		B	-->	C
formation	4	D	-->	E		F

Chaque groupe comprenait 25 personnes qui étaient soumises à un certain nombre de tests linguistiques réceptifs et à un questionnaire contenant un nombre d'échelles d'autocritique. Les tests étaient d'une part des tests globaux, cela veut dire des tests d'aptitude générale (réceptive), de compréhension orale et écrite; d'autre part il y avait des tests qui mettaient à l'épreuve la compétence phonologique, lexicale et grammaticale des étudiants.

Les tests phonologiques, lexicaux et grammaticaux comportaient tous des catégories spécifiques de questions, afin de pouvoir examiner si certains catégories d'éléments ou de règles seraient plus en proie à l'oubli que d'autres. La relation entre la langue maternelle et la langue cible, cela veut dire l'absence ou la présence de contraste entre les éléments ou les

règles des deux langues, était incorporée dans les trois niveaux linguistiques. Les oppositions dans les tests phonologiques étaient en plus divisées en groupe d'oppositions vocaliques et consonantiques; les mots cible dans les tests lexicaux étaient divisés en mots de haute et de basse fréquence.

Les autocritiques se composaient d'autocritiques globales de phonologie, vocabulaire et grammaire, et de ce qu'on appelle les échelles *can-do* pour la comprehension orale et écrite (11 et 8 questions respectivement). Dans le cas des groupes B, C, E et F, les sujets étaient invités à donner non seulement des autocritiques de leur aptitude actuelle, mais aussi de leur aptitude au moment de l'achèvement de l'enseignement du français, des autocritiques dites rétrospectives. Toutes les échelles employées étaient du 'type à cinq points' et tous ces points étaient décrits.

Les groupes qui étaient comparés de façon 'cross-sectionelle' étaient plus ou moins équivalents en ce qui concerne les variables potentielles et pertinentes du passé, telles que la carrière scolaire, les notes de français, le nombre d'années d'enseignement du latin et les contacts parascolaires avec la langue française.

Les résultats de toutes les échelles d'autocritiques indiquaient une perte considérable de la langue selon un modèle remarquablement constant: la quantité (rapportée) de perte était à peu près la même pour chacun des deux niveaux d'enseignement; la perte se manifestait pendant les deux premières années du non-emploi de la langue, tandis que pendant le deuxième intervalle de deux années il n'y avait plus de perte. Les autocritiques globales suggéraient que, selon les sujets, la phonologie était la catégorie la moins oubliée alors que la grammaire était la partie la plus vulnérable à l'oubli.

Cependant les résultats des tests montraient une image tout à fait différente. L'aptitude globale (receptive) demeurait presque inchangée pendant la période de quatre années faisant suite à l'enseignement, et la compréhension orale et écrite progressait même. Les trois niveaux linguistiques montraient des résultats divergents: pour la phonologie il semblait y avoir un progrès, le niveau de vocabulaire restait relativement stable et le niveau de grammaire se dégradait.

Néanmoins l'influence du contraste entre la langue maternelle et la langue cible se sentait pour les trois niveaux. Quant à la phonologie la catégorie 'contraste consonantique' (comme dans les paires minimales

écoute-égoutte et *neuf-neuve*) se trouvait être la catégorie la plus difficile, mais c'était en même temps la catégorie qui contribuait le plus au progrès fait au cours des années. Bien que les scores moyens fussent augmentés, nous avons réussi à identifier un groupe spécifique d'oppositions qui montrait une régression, du moins dans le test écrit de phonologie. De nouveau le contraste y était impliqué, car il s'agissait de l'opposition 'contraste vocalique' entre la voyelle nasale dans, par exemple, *répand* et celle dans *répond*; cette opposition posait de plus en plus de problèmes.

Dans les tests lexicaux, bien qu'il ne fut pas question de perte, le contraste jouait également un rôle important, dans ce cas-là en interaction avec la fréquence. La catégorie des mots qui ne sont pas voisins (*non-cognates*) à basse fréquence, cela veut dire des mots tels que *évier*, *mouche*, *vigne*, *caprice*, et *grève*, se révélait être la catégorie la plus difficile; en plus ces mots-là étaient les seuls mots qui souffraient légèrement d'oubli.

Les tests grammaticaux, finalement, montraient une perte de 10 à 15% de la connaissance préalable, et de nouveau la perte se concentrait sur la catégorie contrastive. En outre, c'étaient surtout les aspects contrastifs du système pronominal français qui étaient sensibles à cette régression.

Le fait que nous ayons trouvé peu de perte peut être expliqué par le fait que nous avions à faire à des étudiants ayant une connaissance relativement élevée du français, surtout en comparaison avec les sujets étudiés dans beaucoup de recherches américaines. En terme de théorie d'apprentissage on pourrait argumenter que tous nos sujets avaient appris le français dans un contexte que les psychologues désigneraient comme '*meaningful learning*'. Et c'est un fait connu que de telles aptitudes sont moins susceptibles d'être oubliées que ce qui est appris par coeur (*rote learning*). En outre, s'il existe quelque chose comme "a critical threshold during learning" ou bien "a critical mass of language", il est tout à fait possible que nos sujets aient en effet atteint un tel niveau d'aptitude du français, qui avait pour conséquence une immunité relative contre l'oubli.

Tout de même il existe de fortes indications que les aptitudes lexicales et surtout les aptitudes grammaticales se dégradent pendant la période examinée. Dans le cas des aptitudes lexicales nous avons trouvé à l'évidence une preuve supplémentaire qui appuie nos théories dans les résultats d'une recherche préparatoire dans laquelle il ne s'agissait pas de mesurer le *succès* mais plutôt la *vitesse* pour retrouver les mots. Cette

évidence suggère que le fait que les tests étaient administrés sans limite de temps - il y avait suffisamment de temps disponible pour les tests oraux, quant aux versions écrites c'étaient les sujets eux-mêmes qui déterminaient l'allure - aurait pu obscurcir la gravité réelle de l'oubli. Ceci pourrait également expliquer pourquoi les sujets étaient tellement pessimistes fâce à leurs aptitudes de français.

Samenvatting (Dutch)

Het onderzoeksproject waarvan hier verslag wordt gedaan had betrekking op de retentie van op-school-geleerd Frans gedurende een periode van twee en vier jaar na beëindiging van het onderwijs Frans waarin er nauwelijks enig contact was met die taal - vandaar de aanduiding 'periode van niet-gebruik'. De proefpersonen waren leerlingen en ex-leerlingen van het VWO die vier of zes jaar Frans hadden gehad. Het gemiddeld aantal uren Frans in dit schooltype is drie per week; met andere woorden, ze hadden 400 respectievelijk 600 uur Frans gehad.

De combinatie van twee onderwijsniveaus en drie tijdsmomenten resulteerde in een onderzoeksopzet met zes groepen. Deze werden in een combinatie van longitudinale en cross-sectionele metingen onderzocht (de pijlen geven longitudinale metingen aan):

		Jaren van niet-gebruik				
		0		2		4
Jaren	6	A		B	-->	C
onderwijs	4	D	-->	E		F

Elke groep bestond uit 25 subjecten, die een aantal *receptieve* taaltoetsen kregen voorgelegd en een vragenlijst die een aantal zelfbeoordelingsschalen bevatte. De toetsen waren, enerzijds, globale toetsen, d.w.z. toetsen van algemene (receptieve) vaardigheid en luister- en leesvaardigheid; anderzijds werden er toetsen afgenomen die de fonologische, lexicale en grammaticale competentie toetsten.

De fonologische, lexicale en grammaticale toetsen bevatten specifieke

sub-categorieën van items, teneinde na te gaan of bepaalde soorten elementen of regels meer onderhevig zouden zijn aan verlies dan andere. Op de drie linguïstische niveaus werd de relatie tussen de moedertaal en de doeltaal verwerkt, d.w.z. de aan- of afwezigheid van contrast tussen de twee talen. De items in de fonologische toetsen waren verder onderverdeeld in vocale en consonantale; de items in de lexicale toetsen in hoog- en laagfrequente.

De zelfbeoordelingen bestonden uit globale zelfbeoordelingen van uitspraak, vocabulaire en grammaticakennis, en zgn. *can-do* schalen voor luister- en leesvaardigheid (11 resp. 8 items). Waar van toepassing - d.w.z. in het geval van de groepen B, C, E en F - werd de proefpersonen niet alleen gevraagd hun *huidige* vaardigheid in te schatten, maar ook hun vaardigheid *op het moment dat zij het onderwijs Frans beëindigden*, zgn. *retrospectieve* zelfbeoordelingen. Alle schalen waren van het vijf-puntstype met alle schaalpunten gemarkeerd.

De groepen die cross-sectioneel vergeleken werden, werden zo goed mogelijk gematched op potentieel relevante achtergrondvariabelen, zoals schoolcarrière, schoolresultaten voor Frans, hoeveelheid onderwijs in het Latijn en buitenschoolse contacten met het Frans.

De resultaten van alle zelfbeoordelingsschalen suggereerden sterk verlies, volgens een opvallend consistent patroon: de hoeveelheid (gerapporteerd) verlies was van ongeveer gelijke grootte voor elk van de twee onderwijsniveaus; het verlies trad op in de eerste twee jaar van niet-gebruik en er was geen verdere achteruitgang in het tweede interval van twee jaar. De globale zelfbeoordelingen suggereerden dat, in de ogen van de proefpersonen zelf, uitspraak het minst achteruit gaat en grammatica het meest.

De toetsresultaten lieten echter een geheel ander beeld zien. De globale (receptieve) vaardigheid bleef vrijwel ongewijzigd gedurende de periode van vier jaar volgend op het onderwijs, en luister- en leesvaardigheid gingen zelfs vooruit. De drie onderzochte linguïstische niveaus lieten uiteenlopende resultaten zien: fonologie bleek vooruit te gaan, vocabulaire bleef relatief stabiel en grammatica ging achteruit.

Desalniettemin kon de invloed van contrast tussen moedertaal en doeltaal worden aangetoond op alle drie de niveaus. Bij de fonologie bleek de categorie 'contrasterend consonantaal' (als in de minimale paren *écoute-égoutte* en *neuf-neuve*) de moeilijkste, maar tegelijkertijd was dit de categorie die het meest bijdroeg aan de vooruitgang over de jaren heen.

Alhoewel de gemiddelde scores vooruitgingen, slaagden we er toch in om een specifieke subset van items te identificeren die enige achteruitgang vertoonde, althans in de geschreven fonologietoets. Opnieuw was contrast in het geding, want het betrof de 'contrasterende vocale' oppositie tussen de nasale klinker in, bijvoorbeeld, *répand* en die in *répond*, die in toenemende mate problematisch werd.

In de lexicale toetsen bleek, alhoewel er geen sprake was van serieuze achteruitgang, contrast eveneens een belangrijke rol te spelen, in dit geval in interactie met frequentie. Laagfrequente niet-cognates, d.w.z. woorden zoals *évier*, *mouche*, *vigne*, *caprice* en *grève*, bleken verreweg de moeilijkste categorie te zijn en bovendien vertoonden deze woorden enig verlies.

De grammaticatoetsen, tenslotte, lieten een verlies zien van 10-15% van de oorspronkelijke kennis, maar opnieuw concentreerde het verlies zich in de contrastieve categorie. Bovendien bleek dat de contrastieve aspecten van het Franse pronominale systeem met name gevoelig waren voor teruggang.

Het feit dat er eigenlijk heel weinig verlies werd geconstateerd, zou verklaard kunnen worden uit het feit dat we te maken hadden met proefpersonen die een relatief grote vaardigheid in het Frans hadden, zeker in vergelijking met de proefpersonen in veel Amerikaanse onderzoeken. In termen van leertheorieën zou men kunnen stellen dat al onze proefpersonen Frans hadden geleerd in een context die psychologen zouden aanduiden als 'betekenisvol leren' ('meaningful learning'). En het is een welbekend feit dat zulke vaardigheden veel minder aan vergeten onderhevig zijn dan van-buiten-geleerde kennis ('rote learning'). Bovendien, als er zoiets is als een 'kritische drempel gedurende het leren' ("critical threshold during learning"), ofwel een 'kritische massa van taal' ("critical mass of language"), dan zou het heel goed denkbaar zijn dat al onze proefpersonen inderdaad een dergelijk niveau van Franse taalvaardigheid bereikt hadden, met als gevolg dat zij relatief immuun waren tegen vergeten.

Niettemin zijn er sterke aanwijzingen dat lexicale en zeker grammaticale vaardigheden onderhevig zijn aan verlies in de onderzochte periode. In het geval van lexicale vaardigheden werd ondersteunende evidentie ontleend aan een vooronderzoek waarin niet zozeer het *succes* als wel de *snelheid* van het ophaalproces werd gemeten. Deze evidentie suggereerde dat het feit dat onze toetsen zonder enige tijdsdruk werden afgenomen - de luistertoetsen werden in een rustig tempo afgewerkt; bij de schriftelijke versies

bepaalden de proefpersonen zelfs *zelf* het tempo - de werkelijke ernst van het verlies enigszins gemaskeerd zou kunnen hebben. Dit zou ook kunnen verklaren waarom de proefpersonen zelf zoveel pessimistischer waren over het onthouden van hun eigen vaardigheden in het Frans.

Rizzumee (Maastricht dialect)

't Ónderzeuk wat hijj gerapporteerd weurt ging euver 't ónthawwe vaan opsjaol-gelierd Frans, twie en veer jaor naotot 't ónderwies waor aofgeslote, zóndertot vääl kóntak mèt de taol had plaotsgevónde - vendao de aonduijing 'periode vaan neet-gebruuk'. De proofpersone hadde veer of zès jaor Frans gehad op VWO-nivo. 't Gemiddeld aontal ore Frans op dat tiepe sjaol is drijj per week, zoetot ze dus rispektievelik 400 en 600 oor Frans hadde gehad. De kombenasie vaan twie trèningsnivoos en drijj tiedsmominte resulteerde in 'nen opzat mèt zès groepe, dee in 'n kombenasie vaan *longitudinaal* en *kross-seksjeneel* metinge woort ónderzeuk (de pijle geve *longitudinaal* metinge aon):

		Jaore vaan neet-gebruuk				
		0		2		4
Jaore	6	A		B	-->	C
ónderwies	4	D	-->	E		F

Eeder groep bestónt oet 25 maan, die 'n aontal *reseptief* taolteste kraoge veurgelag, en 'n vraogelies boe-in ouch 'n aontal zellef-beoordeilinge gevraog woorte. De teste waore, aon d'n eine kant, globaal teste, dat wèlt zègke teste vaan globaal (reseptief) vaardigheid en vaan luuster- en leesvaardigheid; aon d'n aandere kant woorte ouch teste vaan fonologiese, leksikale en grammatikale kinnis aofgenómme.

De fonologiese, leksikaal en grammatikaal teste bevatde allemaol spissefieke soorte vraoge, veur nao te kinne goon of bepaolde soorte illeminte of regels mie ónderhevig zouwe zien aon verluus es aandere. Op alle drijj de lingwistiese nivoos waor de relasie tösse mojerstaol en vreemde taol verwèrrek, dat wèlt zègke de aon- of aofwezigheid vaan kontras tösse de

twie taole. De vraoge in de fonologiese teste waore wijjer ónderverdeilt in vokaal en konsonantaal; de vraoge in de leksikaal teste in hoeg- en liegfrekwent.

De zellef-beoordeilinge bestónte oet globaal zellef-oordeile vaan oetspraok, vokabulaer en grammatika, en *can-do* sjaole veur luuster- en leesvaardigheid (mèt rispektievelik 11 en 8 vraoge). Boe vaan touwpassing - dus bijj de groepe B, C, E en F - woort neet allein gevraog nao d'n tegewoordige kinnis, meh ouch nao de kinnis op 't momint tot ze mèt Frans oetgesjeijd waore, zoegenaomde *retrospektief* zellef-beoordeilinge. Alle gebruukde sjaole waore vief-punts sjoale mèt alle punte gediffinjeerd.

De groepe die kross-seksjeneel vergeleke woorte, woorte zoe good meugelik vergeliekbaar gemaak wat beträf potensjeel relevante achtergrónd-variabele, wie sjaolachtergrónd, sjaolsiefers veur Frans, aontal jaore Letien en boete-sjaolse kontakte mèt 't Frans.

De resultate van al de zellef-beoordeilinge suggereerde sterrek verluus, vollegens 'n opvallend konsistent patroen: de maote vaan verluus waor vaan óngeveer gelieke gruutde veur de twie trèningsnivoos; 't verluus trooj op in de ierste twie jaor vaan neet-gebruuk, en in de daorop vollegende twie waor gei wijjer verluus. De globaal zellef-beoordeilinge gaove aon tot, in de ouge vaan de proofpersone teminste, oetspraok 't wienigste te lijje heet, en grammatika 't mieste.

De tes-oetsleeg daorentege lete 'n gans aander beeld zien. De globaal (reseptief) vaardigheid bleef naogenóg ónveranderd in de veer jaor nao 't ónderwies, en luuster- en leesvaardigheid ginge zellefs veuroet. De drijj lingwistiese nivoos lieverde oeteinloupende resultate op: fonologie bleek veuroet te goon, vokabulaer bleef relatief stabiel, en grammatika ging achteroet.

Neettemin kós 't effek van kontras tösse de mojerstaol en de vreemde taol op alle drijj de nivoos vasgestèld weure. Bijj de fonologie bleek de kategorie 'kontrasterend konsonantaal' (es in *écoute-égoutte* en *neuf-neuve*) 't lestigste, meh tegeliekertied waor dat de kategorie die 't mieste bijjdroog aon de wins euver de jaore heen. Óndaanks tot de oetsleeg veuroet ginge, waor toch 'n spissefiek ónderdeil te vinde wat achteroet ging. Opnuij waor kontras in 't speul, want 't betróf de 'kontrasterende vokaal' oppoziesie tösse de nasaal klinkers in, beveurbeeld, *répand* en dee in *répond*, die in touwnummende maote problematies weurt.

In de leksikaal teste späölde kontras ouch weer 'n belangrieke rol, meh

noe in interaksie mèt frekwensie. De lieg-frekwente, neet-verwante wäörd ('non-cognates') wie *évier*, *mouche*, *vigne*, *caprice*, en *grève*, bleke verreweg 't meujelikste, en dat waore ouch de wäörd die get achteroet ginge.

De grammatika-teste tenslotte lete zien tot 10-15% vaan d'n oorsprónkelijke kinnis eweg waor, meh 't verluus konsentreerde z'ch in de kontrastief kategorie. Wijjer bleke de kontrastief ónderdeile vaan 't Frans pronominaal systeem bijj oetstek geveulig veur trökgaank.

Tot erreg wienig verluus gekónstateerd woort, zouw verklaord kinne weure door 't feit tot v'r mèt proofpersone gewèrrek höbbe die 'n erreg hoeg nivo vaan Frans hadde, zeker in vergelieking mèt de proofpersone in vääl Amerikaans ónderzeuk. Lier-teoreties besjouwd zouw me kinne zègke tot alle proofpersone Frans hadde gelierd in 'ne kónteks dee psychologe zouwe karakterisere es 'beteikenisvol liere' ('meaningful learning'). En 't is algemein bekind tot beteikenisvol liere vääl minder ónderhevig is aon vergete es vaan boete liere ('rote learning'). Dao kump nog bijj, es zoeget besteit wie 'ne 'kritieken dörrepel bijj 't liere' ("critical threshold during learning"), of 'n 'kritieke massa vaan taol' ("critical mass of language"), tot 't meugelik is tot al us proofpersone inderdaad zoe'n nivo in hun Frans bereik hadde, mèt es gevolleg tot ze relatief immuun waore tege vergete.

Neettemin höbbe v'r sterreke aonwijzinge gevónde tot leksikale en zeker grammatikale kinnis achteroet geit in de periode die v'r ónderzeuk höbbe. Wat de leksikale kinnis betröf, daoveur kóste v'r aonvöllend bewies aonvoere oet 'n veurónderzeuk boe-in neet zoezier 't *suuksès* vaan 't ophaolproses, es wel de *snelheid* devaan, ónderzeuk woort. 't Feit tot us teste zónder einigen tiedsdrök woorte aofgenómme, zouw bès wel ins de woeren errens vaan 't verluus gemaskeerd kinne höbbe. Dat zouw ouch kinne verklaore boeveur de proofpersone zellef zoe pessimisties waore euver 't ónthawwe vaan hun Frans.

Curriculum Vitae

Bert Weltens werd geboren in 1957 in Heer-Maastricht. Na het behalen van het Gymnasium β diploma ging hij in 1975 Engelse Taal- en Letterkunde studeren aan de Katholieke Universiteit te Nijmegen. In 1978 deed hij daarin Kandidaatsexamen, waarna hij vervolgens Algemene Taalwetenschap ging studeren. Zijn doctoraalexamen behaalde hij in 1981.

Na zijn afstuderen verwierf hij een *British Council Scholarship*, dat hem in staat stelde een jaar aan de University of Reading, Engeland, te studeren en, onder begeleiding van Peter Trudgill, sociolinguïstisch onderzoek te doen. Na zijn terugkeer in Nederland in 1982 werd hij aangesteld als projectuitvoerder bij het Instituut voor Fonetiek van de K.U.N., bij het project *Visuele terugkoppeling van Nederlandse intonatie.*

In februari 1984 begon hij aan het Instituut voor Toegepaste Taalkunde met het NWO-project waarover in dit proefschrift wordt gerapporteerd. Het onderzoek liep tot september 1988, maar naast zijn aanstelling bij NWO werd hij met ingang van september 1987 voor de duur van twee jaar benoemd tot part-time universitair docent aan hetzelfde instituut.

STELLINGEN

behorende bij het proefschrift
The attrition of French as a foreign language
te verdedigen op 14 december 1988 door H.H.G. Weltens

1. De wijdverbreide overtuiging dat vreemde-talenkennis heel snel weer wegebt als je haar niet regelmatig gebruikt, is gezien de resultaten van deze dissertatie op zijn minst als ongenuanceerd te kwalificeren.

2. De pessimistische kijk van de proefpersonen zelf op hun retentie van het Frans, zoals die blijkt uit hun zelfbeoordelingen, hangt ongetwijfeld samen met de mythe dat Frans zo'n moeilijke taal is: niet alleen moeilijk om te leren, maar ook om te onthouden.

3. De roep om interne-consistentie betrouwbaarheid bij het zien van somscores is lang niet altijd een uiting van gefundeerde zorg over de *error*-bestendigheid van conclusies.

4. Het aanhoudende geweeklaag over de teloorgang van de zo geroemde vreemde-talenkennis van de Nederlander lijkt in ieder geval indruk te hebben gemaakt op de Staatssecretaris voor Onderwijs en Wetenschappen, nu zij zo nadrukkelijk pleit voor het verplicht stellen van een tweede vreemde taal bij de HAVO/VWO-eindexamens.

5. De verplichting om in een Nederlandstalige dissertatie een Engelstalige samenvatting op te nemen heeft alleen zin als er ook eisen worden gesteld aan de begrijpelijkheid van die samenvatting voor Engelssprekende lezers.

6. De totstandkoming van dit proefschrift is niet bevorderd door het instellen van "zwaartepunten", "aandachtsgebieden", "speciale programma's", "kamers", "netwerken", "expertisecentra" en "persoonsgerichte groepssteun".

7. Gezien de grote rol die de popmuziek speelt in de 'incidentele' verwerving van vreemde-talenkennis, verdient de positie van het *Brits*-Engels als norm in het vreemde-talenonderwijs ernstige heroverweging.

   (*zie* Peter Trudgill (1983), Acts of conflicting identity The sociolinguistics of British pop-song pronunciation In *On dialect. Social and geographical perspectives* Oxford Blackwell, 141-160)

8. Enerzijds maakt het ontbreken van een gecodificeerde norm dialecten gevoeliger voor veranderingsprocessen; anderzijds ontlenen zij juist daaraan een belangrijk deel van hun vrijheid en vitaliteit.

9. Als de behoefte aan intermenselijk contact de basis is voor taalverwerving, dan begint de verwerving van de moedertaal niet bij de geboorte, maar ruimschoots daarvóór.

   (*zie* Thomas R Verny & John Kelly (1981), *The secret life of the unborn child* New York Summit Books)

10. In tegenstelling tot wat bijvoorbeeld de Amerikaanse, Zweedse en Westduitse pers doen voorkomen, zijn er objectieve argumenten aan te voeren vóór het Nederlandse drugsbeleid.

11. Zolang de regelingen voor kinderopvang in Nederland zo gebrekkig blijven, zouden ze méér ten koste moeten gaan van de arbeidsproduktiviteit van jonge *vaders*.

12. Het hardnekkige gebruik van "Holland" als aanduiding voor ons land zou bestreden moeten worden, in ieder geval op vrachtwagens afkomstig uit Nederlands-Limburg.